AFRL-VS-PS-
TR-2004-1206

AFRL-VS-PS-
TR-2004-1206

RECONFIGURABLE WIRES

James Lyke

31 December 2004

Final Report

AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
3550 Aberdeen Ave SE
AIR FORCE MATERIEL COMMAND
KIRTLAND AIR FORCE BASE, NM 87117-5776

DTIC COPY

AFRL-VS-PS-TR-2004-1206

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report has been reviewed by the Public Affairs Office and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nationals.

If you change your address, wish to be removed from this mailing list, or your organization no longer employs the addressee, please notify AFRL/VSSE, 3550 Aberdeen Ave SE, Kirtland AFB, NM 87117-5776.

Do not return copies of this report unless contractual obligations or notice on a specific document requires its return.

This report has been approved for publication.

//signed//

JAMES LYKE
Project Manager

//signed//

KIRT S. MOSER, DR-IV
Chief, Spacecraft Technology Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
31-12-2004	Final Report	01-06-2002 to 31-12-2004

4. TITLE AND SUBTITLE
Reconfigurable Wires

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
65502F

6. AUTHOR(S)
James Lyke

5d. PROJECT NUMBER
2311

5e. TASK NUMBER
RP

5f. WORK UNIT NUMBER
ZZ

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
3550 Aberdeen Ave., SE
Kirtland AFB, NM 87117-5776

8. PERFORMING ORGANIZATION REPORT NUMBER

AFRL-VS-PS-TR-2004-1206

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The "Reconfigurable Wires" project explores the reliability of gold-gold surfaces used in some microelectromechanical systems (MEMS) implementations of relays and postulates alternate formulations. The research has the exciting potential to: (1) immediate improvement in reliability of MEMS relays; (2) development of new devices that operate in non-traditional current / voltage regimes; (3) improvement in the current handling, constriction resistance, and reliability of macro-relays.

The three primary products of this research to date are: (1) a journal paper on MEMS nanoscale contact surface morphology, (2) an invention disclosure on arc suppression schemes in micro-relays, and (3) an unpublished finding on the properties of 2-D randomized arrangements of "sticks" in 3-D.

15. SUBJECT TERMS
MEMS, reconfigurable wires

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. James Lyke
Unclassified	Unclassified	Unclassified	SAR	82	**19b. TELEPHONE NUMBER** *(include area code)* (505) 846-5812

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

1.0 Introduction

Reconfigurability has become very important to 21st century systems. This report addresses reconfigurable systems with an emphasis on wires. Wires are important as they related to the interconnections between components. Take a collection of components and throw them on a table, and you have a bunch of components. Connect the same components together, and you have architecture. Being able to recompose the same components into different architectures under software control gives you a reconfigurable architecture. The most well known real-world example is the field programmable gate array (FPGA). It represents a sort of "digital parts cabinet" in which a user, through a software program, can define the connective relationship between logic and memory to form a digital system. It is partly an illusion, because the physical structure is pre-defined, and the FPGA is programmed only by re-arranging charge configurations in this pre-built piece of silicon. It is in fact necessary to pre-build a superset of wiring possibilities, a subset of which is selected for any one user design. If an illusion (or more correctly an emulation) the illusion is becoming more effective as driven by Moore's law. At the time of this writing, systems with more than 10 million effective gates can be defined this way.

Reconfigurability 101

What it reconfigurability? One definition, perhaps the most trivial, is that reconfigurability is the ability to define multiple settings of function or structure in a deployed system on command. In other words, it is about putting lots of "knobs" in a system that can be turned under software control. The knobs can be nano (within an integrated circuit) or mega (platform scale).

Does this simple definition imply that everything is reconfigurable to some level? Yes. Are all reconfigurable systems equal? No, or in other words all manifestations of reconfigurability are not equal and may not be useful. Consider the example of mating a legacy transceiver to a new reconfigurable computer containing FPGA devices. If the transceiver was built in the 1970's, chances are good that its digital interface might have TTL-compatible (5.0 volt) signals. Modern FPGA interfaces support many different interfaces, but not at 5.0 volts. So all of the configurability possible in that FPGA will not effect a simple capability to connect together those two boxes. If, instead, the reconfigurable FPGA devices in the new computer were connected to a programmable level shifter which could be adjusted in software to span a 5.0 volt interface, then it might be possible to connect the boxes together. But maybe then the connector pins are in the wrong order. The reconfigurability of the level shifter and the FPGAs combined might be able to correct this problem, unless one of the mismatched conductors is a primary power line. If, however, we now add a programmable wiring harness between the two boxes, then we may be easily able to salvage the situation, by re-configuring the wiring patterns between the two boxes. This simple example required three distinct forms of reconfigurability.

In the very simple definition, one might ask "what is really new"? Certainly, as suggested by the previous examples, part of the difference in modern reconfigurability approaches is a systematic organization of reconfigurability into domains. Reconfigurability can be viewed as having taxonomy. Taxonomies lead to disciplines. Early software development was less systematic and less disciplined before the tenets of "structured programming" were implemented. In the earliest computing systems, the idea of self-modifying code may not have appeared improper, but even might have been considered "clever". Only after the nightmare of debugging self-modifying code might it have occurred to systems designers that self-modifying code was a bad idea. Software design eventually evolved into a systematic versus ad hoc practice. Systematic approaches lend themselves to rigorous study and automation. So, too, is it likely that reconfigurable systems will undergo an evolution into systematic approaches to harness the "knobs" in a disciplined way.

The Organization of This Report

The remainder of this report will explore a number of facets of reconfigurability in systems and devices and will examine the limit arguments for interconnections in nanoscale systems. Chapter 2 extends the "reconfigurability 101" discussion considerably, detailing the taxonomy of reconfigurable systems in more detail. Chapter 3 focuses on one reconfigurable systems technology, the programmable wiring harness. Chapter 4 "drills down" further on the study of reliability in microelectromechanical systems (MEMS) switches, as such switches are of keen interest in the programmable wiring harness. Finally, Chapter 5 examines the statistical properties of nanoscale wiring systems based on randomized three-dimensional distributions. This examination is a sort of limit argument to molecular

scale interconnection schemes and demonstrates an interesting correlation between randomness and scale-free wirelength models.

2.0 Reconfigurability in Systems

Introduction

Embedding flexibility into the design of fielded systems is a trend important for the future. Flexibility, among other things, provides for the possibility that a system might adapt to unforeseen events, to continue in support of its originally defined mission or to accommodate a new mission concept altogether. The change may be as seemingly insignificant as the correction of a mathematical constant in a thrust algorithm, or as radical as changing a battle tank into a gunboat. The limiting factor is technology, and it is generally appreciated that the technology for altering bit patterns in a register exists today but not for altering the patterns of material in a structure. However, particularly within the domain of electronics, the "range of influence" of technologies that can be reconfigured under software control has expanded and continues to grow, increasing the "space" of reconfigurable features that could be directly embedded in systems.

It is this expansion in reconfigurable technologies that is the focus of this paper, especially as it may apply to space systems. The next section discusses field programmable gate arrays and their relative state-of-the-art (SOTA) in space systems. Section III discusses the motivation and taxonomy of a generalized reconfigurable electronic system. The following section examines strategies for how a system with reconfigurable characteristics might be harnessed. Section 5 describes a roadmap for reconfigurable systems for space.

Reconfigurable Gate Arrays

In considering the larger picture of how reconfigurable systems architectures for space might emerge in the future, it is instructive to examine the evolution of field programmable gate arrays (FPGAs) in both terrestrial and space systems. Before the advent of programmable logic devices, the development of digital systems required the use of a number of discrete integrated circuits or the design and fabrication of custom application-specific integrated circuits (ASICs).

In electronics, architecture is the functional expression of a connective relationship between the electrical terminals of devices at some scale of regard, usually an entire integrated circuit. Reconfigurable architectures add a level of indirection, by providing an illusion of underlying structure that in fact can be altered under program control. The emergence of FPGAs provided within the confines of a monolithic integrated circuit (IC) the capability to define a seemingly arbitrary connection of logic, memory, and interconnection resources as required to implement complex Boolean circuits. Rather than resort to a collection of discrete, elemental ICs or a fully integrated monolithic IC (requiring large amounts of time and money to produce a single prototype), it was possible for the first time to "impress" a user's design within a pre-fabricated component. This capability became very important, primarily because time and expense required to produce prototypes using FPGAs is minimal compared to a custom-fabricated IC. Even when a design could be formed from a collection of discrete parts, the use of a FPGA in some cases was preferred, since changes in the terminal relationships within a FPGA could be formed without the need to rewire a circuit board, as would often be required in a design based on many discrete parts.

FPGA technologies can be defined in two broad classes: one-time programmable and reconfigurable. One-time programmable FPGAs are commonly based on an antifuse approach in which normally open paths in interconnections can be irreversibly shorted. While such FPGAs offer high performance, they offer little prospects for circuit alteration after programming. The most flexible FPGA devices are reconfigurable, usually employing SRAM configuration cells, which can be reprogrammed repetitively.

Reconfigurable FPGAs are differentiated then by ASICs in that otherwise fixed logic functions and interconection patterns are replaced by programmable logic cells and programmble routing networks. Programming is achieved at a price, since programmable resources require more physical real estate (as much as 100X compared to a fixed technology [1]). The overhead comes about from three primary factors:

- the use of lookup tables (LUTs) to implement logic functions;
- the use of an interconnection network in which many transistor switches are embedded to permit complex configurations of internal connectivity;
- the use of a static storage cell to maintain the state of every LUT and interconnect switch, as well as other configuration options within internal and input/output cell structures.

It is important to note that SRAM-based FPGAs employ static memory cells in two fundamentally different ways. *User memory*, for example, refers to flip-flops, registers, and other memory structures deliberately invoked by users as part of a target Boolean design. *Configuration memory*, on the other hand, refers to memory bits normally inaccessible to users. These memory structures are used within LUTs, programmable routing, and other parts of the FPGA as required to realize a user design. Since the upset of memory elements in on-orbit systems is of great concern, FPGAs have seen even more limited acceptance for space use than other types of commercial parts. If, for example, the user memory of a design is affected by radiation, then it is arguably possible to cope with the upset using some sort of error detection and correction approach applied to the data. Even entire circuits in a digital design can be replicated. However, if configuration memory is upset due to a radiation-induced single event upset, the results are far less manageable since the circuit topology itself is modified and simplistic voting or redundancy strategies are not directly applicable.

These concerns did in no way impact the spectacular success of these devices in "terrestrial" settings. Early SRAM FPGA devices had a limited effective gate capacity (dozens to hundreds of gates), and the programmable structures were of such limited flexibility as to require an intimate knowledge of the FPGA's internal architecture to exploit them in the few cases where these early devices could be applied. The power and flexibility of FPGAs, however, grew substantially to the present, with present-day devices theoretically capable of expressing designs in excess of one million gates. The scale of integration in modern FPGAs is so large that comprehensive hand-tuning is out of the question and computer-aided design (CAD) flows similar to those used in ASICs must be employed, starting with design capture based on synthesizable descriptive languages, such as VHDL and Verilog. While the overhead factors previously cited are still present, the density and performance of modern IC processes provide levels of density and speed, even with such handicaps, far beyond that of the full-custom ICs built just a few process generations before. With the exception of "commodity" components, such as standard processors and memories, an increasing fraction of new designs start in FPGAs and remain there, unless raw performance and/or production scales can provide an economical motivation to migrate to custom silicon.

Aerospace systems, though stereotyped as having the highest performance, receive tremendous benefits from the use of FPGA devices. Since the average build of a DoD research system is very low with few exceptions, FPGAs dramatically improve the economy of R&D. While this economy is true for one-time programmable FPGAs, the benefits for using reconfigurable FPGAs with in-system reprogramming can be dramatic. Complex systems that employ FPGAs not only enjoy the ability to incorporate late-point design adjustments (DoD systems have a notorious propensity for "requirements creep"), but in-system reprogrammation of reconfigurable FPGAs can be a tremendous asset in test and debug during critical stages in development. The prospect of building large, complex custom ASICs is analogous to creating very complex software programs that can be changed only once every four-six months at a tremendous expense. Since even good software code is far from defect-free, the development of custom silicon requires an unprecedented level of confidence in requirements, design, and verification. Even in the cases where custom silicon must be used in a system design (usually for added performance), it is becoming more common to see FPGA brassboard precursor systems employed preceding the commitment of a design to custom silicon for the added confidence a of functional hardware implementation beforehand.

Space systems electronics do not enjoy the availability of reconfigurable FPGAs. This is perhaps surprising, when the introduction of even elementary one-time programmable FPGAs to space has been enthusiastically received. The first AFRL (formally Phillips Laboratory) program to develop an FPGA was based on establishing radiation-hardened versions of the Actel 1020 and Actel 1280 devices (~1K and ~10K effective gate capacities, respectively) [2]. Even with significant limitations, such as susceptibility to single event upset (SEU) in user memories, these Actel components were among the most successful components developed in radiation-hardened form, and they are used widely in space systems designs today.

In the 1990's, two abortive attempts to create radiation-hardened reconfigurable gate arrays (based on Xilinx XC3000 series FPGAs) were undertaken. The first attempt, initiated by Intersil (formerly Harris in Melbourne FL), would have resulted in a radiation-hardened XC3020 (2K effective gate capacity), but was terminated due to the de-emphasis by Harris on high-end digital radiation-hardened microlectronics in the mid-1990's. More recently, Phillips Laboratory initiated a second program with System and Process Engineering Corporation (Austin TX) [3] to make a radiation-hardened version of the XC3090/XC3190 FPGA in complementary heterostructure FET technology, which would have resulted in an extremely fast FPGA relative to silicon devices at that time. This

effort, however, also fell victim to business case issues relative to the CHFET technology, resulting in the eventual termination of the program.

More recently, at least two efforts are endeavoring to make radiation-hardened reconfigurable gate arrays based on alternate approaches. In the first effort, a NASA-funded effort with Honeywell has been pursuing the radiation-hardening of an SRAM-based FPGA based on the Atmel AT6010 (30K gate equivalent) architecture [4,5]. In this program, the Atmel architecture is re-cast in a Honeywell radiation-hardened process In the second effort, a non-commercial architecture has been developed "from the ground up" *as* a radiation-hardened FPGA [6]. This architecture, referred to as "Orion", was based originally on a concept to build an FPGA around the Northrup Grumman SONOS non-volatile memory technology [7]. The Orion series has since evolved to a SRAM-based technology, based currently on two planned offerings, a 4K gate equivalent device and 64K gate equivalent device, planned for availability by about 2003. The Orion and presumably the Honeywell architectures are developed with special resilience to SEU. Such structures for FPGA configuration memory cells have been recently reported [8]. MRC's FPGA, for example, uses dice latches and other structures to mitigate single event transient effects [9], perhaps the first FPGA concept to offer this additional level of robustness. The FPGA devices are otherwise designed to be latch-up free and tolerant to moderately high levels of total ionizing radiation.

If a high tolerance to single event upset represents the minimum acceptable threshold for FPGAs to be used in space, a number of researchers are aggressively pursuing sub-standard approaches. A number of users, either unwilling to await the emergence of more robust FPGAs or find their relatively modest gate capacities unacceptably low, are attempting to forge constructions of FPGAs that can be used in certain types of space applications. Subjecting SRAM-based FPGAs to radiation and/or designing space electronics around them appears to be an internationally popular practice, based on work reported by authors from not only the US [10,11], but also Australia [12, 13], Canada [14, 15], France [16], Italy [17], and PRC [18]. Most of these examinations are based on Xilinx parts, particularly the higher gate capacity Virtex series components, with only one reported investigation of an Altera device (10K100A, a 100K gate equivalent component) [17]. Among the many debatable assertions are that Xilinx optimized the devices for use in space [15], that the relative propensity to upset is very small, and that the latency to detect and completely reconfigure an entire FPGA (which can take several hundred milliseconds) is generally acceptable [19]. Less debatable is the compelling motivation *to* use such devices in space. The strategies for using upsettable FPGAs in space hinges on two fundamental requirements: (1) that the FPGA be immune to destructive latch-up and (2) that the FPGA have some mechanism for verifying the integrity of the configuration memory bitstream. If these two requirements can be met, then it is possible to use SRAM-based FPGAs in space applications, although disruptions in the configuration and data memory are possible. In the case of the configuration memory, it is possible in some Xilinx components to output configuration memory for purposes of comparison to a known valid copy of the bitstream. Any differences can trigger a complete bitstream "refreshment." User memory, however, has no such systematic mechanism for upset detection, and users are forced to rely on other practices, such as triple modular redundancy, to ensure integrity in the data storage elements of a user application.

Taxonomy / Motivation For Reconfigurable Systems

For the purposes of this paper, it may be sufficient to define "reconfigurable" as "software-accessible". In particular, we may regard that reconfiguration makes available a series of "knobs" or adjustments that can be controlled through information computation and communication processes. Albeit in an FPGA, the number of such available adjustments is very large (millions). In this case, a user is provided with at least the illusion that great quantities of logic gates, memory elements, and interconnect resources are available to be freely manipulated within the confines of a fielded embedded system, even under remote control. This latter possibility of altering the arrangement of parts of a system is powerful and sometimes frightening. If conditions can be established to effect desired manipulations to a pool of reconfigurable resources and at the same time prevent undesired, pathological adjustments, then the effect can be particularly enabling. Under the assurance of these conditions, it is furthermore desirable to open up a larger space of consideration, since computers and FPGAs as we usually regard them are purely binary, discrete systems and most of the "real world" is not. In this section, we propose a taxonomy for reconfigurable systems that exposes many other types of "knobs". It is further possible to discuss singular or collective versions of reconfigurable systems

Taxonomy

A possible taxonomy or hierarchy for a reconfigurable system is presented in Figure 1. As a hierarchy, the proposed set relationships and containments are not rigorous, but intuitively approximate. The innermost or lowest level containment is the digital regime, which includes traditional and reconfigurable computers along with FPGAs. The set relationship within digital suggests an overlap between computers and FPGAs as opposed to the latter containing the former. The analog regime does consider all digital as being a subset, since digital systems in practice involve discretely interpreted analog signal patterns that are manipulated by analog components that are considered digital by virtue of a logical system based on input and output decision thresholds. The analog regime, however, is far more broad, as it includes high-frequency signaling (the demarcation of which is considered to be delimiting of lumped element and transmission line behaviors in the electrical signals), field programmable analog components and circuits, and configurable power drive (dc or ac) electronics. An overall *electrical* regime, referred to as "programmable wiring", is shown to contain all previously discussed forms of reconfiguration, since electronics requires "electron pathways", a principle role of wire. With wires defined by software, such pathways can be reconfigured. The final and perhaps most controversial regime suggests a further over-arching regime, that of *programmable matter*. Occupying this tier are items that might rightly be placed into a category of science fiction, but could also contain a large set of more common elements, such as actuators or configurable structures.

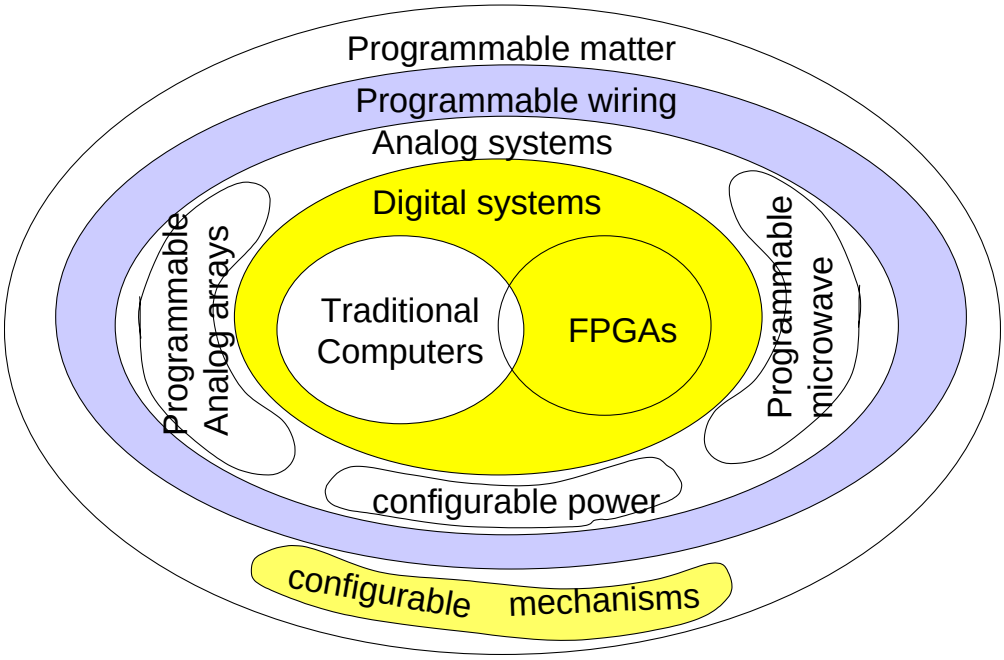


Figure 1. An attempt to establish order within a reconfigurable universe. This diagram suggests a possible taxonomy for reconfigurable "mono-systems".

Networks of Reconfigurable Object - Extension of the taxonomy concepts to networks is likewise proposed. In order to be considered a network, a collection of reconfigurable objects must have the ability to co-mingle information and control, and the reconfigurable properties of network elements are exposed to the network for control via an applicable reconfiguration process. A singular object with reconfigurable properties could be referred to as a reconfigurable *mono-system*. A collection of reconfigurable objects with "short-range order" could be referred to as a reconfigurable *cluster* or reconfigurable *plasma*. The definition of "short-range" is relative, but consistent with the perception that the cluster in some sense appears to take on the character of a single object.

Finally, a reconfigurable *constellation* refers to any other type of structured network, whereas any other general amorphous network is simply referred to as a reconfigurable network.

Reconfiguration Processes and Methods - The ability for a feature to be reconfigured does not necessarily imply a specific process for change. In other words, the existence of a "knob" is a separate concept from how the knob is "turned". If a client is defined as a reconfigurable entity, then the process of "knob management" must be considered. Managing "knobs" requires the existence of a configuration management function or subsystem. This configuration management processor (CMP) can be a simple state machine or auxiliary computer that establishes order and oversight of the reconfigurable features of a client system. Obviously, the basic role of a CMP is to introduce changes in a controlled, predictable manner. Secondary roles of the CMP is to observe the state of a client and to manage the abstraction (for example, as a graph structure) the represents the reconfigurable "resource pool". The latter function becomes particularly important in systems that may undergo expansion or reduction due to subsystem additions, deletions, or other modifications that might result from defects or failures over time. A single *system can employ several CMPs, which themselves can be networked together. A single CMP serves the role of master among a pool of CMPs, and the others are referred to as slaves.*

Recognizing the need for a CMP in a reconfigurable system, it is now possible to discuss a number of dimensions around which CMP operation might be organized.

Frequency of change - Essentially for basic temporal categories can be defined based on how frequently a client needs to be reconfigured:

- *One-time.* In this case, the reconfigurable portions of a system are defined "at the factory" and are not expected to be altered in the future. This case is demonstrated in anti-fuse FPGAs, fusible PROMs, and other firmware concepts. It is useful in the case where reconfigurability is used primarily to rapidly customize and build a variety of systems using a limited inventory of flexible components, but where no further expectations are placed on the need to alter those configurations.
- *Static.* In this case, the ability to perform in-system reconfiguration is useful, but it is not necessary to adjust the overall system configuration during a mission "session". This accommodates the equivalent of a reconfigurable system "service package upgrade", to permit the possibility of fine-tuning or "bug-fixing" of a post-deployed system. This approach has been used on programs where a mechanism for on-orbit software uploads have been accommodated.
- *Quasi-static*. In this case, the need to alter system configuration is more frequent and incompatible with an approach of sending software upgrades each time a change is required. In this case, the system must somehow bank its various *personalities*, where a "personality" is defined as a distinct configuration description, context, or bitstream. Reconfiguration is prompted in response to system events or mode changes. Since the time for in-system reconfiguration may be long relative to the time scale of events and certainly longer than individual computation cycles, this method is useful only when the gains of reconfiguration warrant placing those portions of a system into an offline condition while changes are being made. In other words, if the period between reconfigurations is a computation *epoch*, then in quasi-static reconfiguration, the time between computational epochs is usually longer than the time necessary to reconfigure [1].
- *Dynamic.* In dynamic reconfiguration, the time to reconfigure is comparable to the scale of individual computing events. This domain, sometimes referred to as *context switching*, is "agile" in the sense that reconfiguration must be rapid to be meaningful.

Degree of change - Reconfiguration can be total, or partial, depending on the ability of components to support isolation of particular segments for reconfiguration. In order to support dynamic reconfiguration without partial reconfiguration, it is necessary for each reconfigurable resource to maintain locally enough memory to locally store each context and to permit the rapid selection between these contexts by multiplexing. Partial configuration is desirable, especially for bandwidth-limited systems, as it permits reconfiguration with smaller bitstreams.

How and where change is computed and transferred - The most basic method for developing the bitstreams that define settings in a reconfigurable system involves careful offline scripting and compilation. This process is analogous or identical to writing computer software in a high order language. The compilation process produces a bitstream or object code, which can be transferred into a reconfigurable system from an external source. This approach can be satisfactory in quasi-static and dynamic reconfiguration so long as the number of personalities do

not exceed on-board storage and do not need to be replenished faster than they can be regenerated externally and uploaded. This situation can be referred to as a "reconfigure-from-pallete" approach, in which the CMP serves as a "jukebox" or personality server to the reconfigurable system. When, however, the need for personalities involves an impractically large combinatorial set and waiting for offline computation/upload is impractical, then it is to consider approaches for compiling personalities organically, i.e., within the reconfigurable system itself. In this *self-configuration* approach, the CMP itself compiles bitstreams as required for internal banking and/or direct transferal into the reconfigurable resources of a system. Self-configuration is a powerful concept in that it permits a system to possess "reactive" interfaces. Reactive interfaces can be thought of as an advanced form of plug-and-play, in which a system is reconfigured in response to a particular combination of components that are plugged into it, and the compilation process is done by the system itself. A full discussion of reactive interfaces is beyond the scope of the present paper, but some characteristics of the approach are briefly discussed. First, the process of interface identification requires that some procedure be established to clearly specify signal, control, and power supply requirements of components that participate in a reactive interface. In general, the specification itself might be done with an existing or special-purpose scripting language. The reactive interface would transfer these electronically encoded interface requirements to the CMP through some standard interface, possibly in the form of dedicated pins in reactive interface connectors. The scheme is analogous to the "Transducer Electronic Data Sheets" used in the IEEE1451.2 smart sensor standard [20]. The CMP would merge all scripts involved and compile a "production" bitstream. This bitstream could be transferred directly into the reconfigurable resources of a system and banked in CMP memory. Checksums could be appended to the stores for reducing the time of subsequent reconfiguration, since compilation processes are typically time-intensive. The notion of partial recompilation or more dynamic on-line bitstream preparation techniques can be considered, but shall not be further discussed here.

Summary of Motivation for Using Reconfigurable Systems

As the definition of a taxonomy appears to be a preliminary step in defining a science or at least discipline of reconfigurable systems, it is once again germane to consider the question of why reconfigurable systems are important. Some of these reasons were introduced already in the preceding discussion on programmable gate arrays and in the opening paragraphs of this paper. Here we attempt to expand and reinforce those arguments, especially as they pertain to space systems:

Improved fault tolerance—Reconfigurability offers a greater repertoire of techniques for responding to system faults, and in principle the more levels of reconfigurable applied to a system increases the possibility for recovering against a larger range of failure modes. For example, if large number of "software wires" are included in a system wiring harness, bus failures might be circumlocuted.

Optimization of performance—Given enough "knobs" and the flexibility to refine arrangements of components, an otherwise fixed resource pool could in principle be re-optimized to respond to new requirements.

Energy management—As a corollary, many unneeded parts of a system could be de-energized in a much more fine-grain manner than currently possible.

Reduced system development time and cost—The elimination of the time and expense for one or more customized fabrication cycles is possible through the use of reconfigurable components. Furthermore, reconfigurability can permit improved testing a configuration refinement through improved controllability and perhaps observability. Specialized formats of reconfiguration optimized for troubleshooting that cannot be considered in fixed systems can be developed, theoretically reducing the amount of effort required to expose latent design.

Improved ability to meet performance requirements— Thomke [21] established empirical evidence of a causal relationship between flexibility and "development performance" (the degree to which a developed system meets customer requirements). Typically, requirements change in any complex system, according to Thomke, is due to (1) co-evolution of component technology and (2) the inherent inability of users to accurately specify their needs at the outset of a development effort.

Adaptiveness to Improve Mission Response—Flexibility by virtue of in-system reconfigurability allows platforms to be optimized for a current mission, perhaps to cope with boundary conditions unknown or uncertain at the time of

development. More importantly, the ability to reconfigure permits to possibility that a fielded system's might adapt to perform new missions not originally planned.

A Reconfigurable Systems Roadmap

This section outlines strategies for reconfigurable components, subsystems, and systems based on the proposed taxonomical concepts. Some of the elements introduced here will be recognized as fundamental, while others must be considered propositional. A possible reconfigurable mono-system framework is also described and a number of challenges in hardware, software, and resource management are discussed.

Reconfigurable Component Strategies

Reconfigurable systems will require fundamental building blocks with properties that can be manipulated.

Switches—Switches provide fundamental electronic pathway control and are the most prevalent function of transistors in modern electronics. Transistor-based switches will remain the most important single device in reconfigurable systems, but other technologies will be required, since transistors are limited to applications where: (1) voltage ranges do not destroy or alter the device operation; (2) energy supply availability is important. Latching switches based on micro-electromechanical systems (MEMS) technology are attractive as a supplement to transistors, particularly when "software wires" are needed whose states can be maintained in the absence of power application. Unlike the much larger conventional relay counterparts, MEMS switches can be integrated at densities above 100 devices/cm^2 [22].

Non-volatile Storage—Non-volatile memory (NVM) is abundantly required to preserve the state settings of the many forms of "knobs" in reconfigurable systems. NVM can reside in coarse-grained storage banks and/or be finely distributed throughout the a system. NVM in space has been particularly problematic, and gains in the density of radiation-hardened media have been slow. Existing technologies offer poor levels of integration (density scale < 1Mbit/device), and the challenge remains one of the most important for future space electronics. The current front-running approaches include magneto-resistive memory, SONOS, and chalcogenide ($Ge_2Sb_2Te_5$) memory. The latter approach, which can be crudely viewed as a programmable resistor, is seen as particularly promising [23] due to the relatively high maturity of the associated thin-film deposition, high-cycle reliability (>10^{13}), dense integration (single-transistor cell), multi-bit storage/cell, and good radiation tolerance [24].

Field programmable gate arrays and interconnect devices (FPGAs/FPIDs) — All digital systems are comprised of logic, memory, and interconnection resources. Reconfigurable digital systems, such as the FPGA, permit the arrangements of otherwise fixed resources to be defined by a user. From the standpoint of theoretical computer science, reconfigurable processing did not transcend the Turing machine model, but instead provided the possibility of recasting given computing problems in entirely different formats. This recasting can have significant practical benefits. Since any computation can be viewed as having a time-space complexity product, reconfigurable processing permits a more detailed elastic trade between temporal and spatial computation in scheduling arrangements of hardware resources to solve a problem at hand. In many cases, this leads to a dramatic speed-up due to hardware parallelism. The geometrically increasing popularity of FPGAs (from a niche market to a multi-billion dollar industry) testifies to the potential market for at least this one class of reconfigurable system.

The FPGA and its interconnect-only “cousin”, the field programmable interconnect device (FPID), became the heart of many concepts for reconfigurable computing in the 1990s. Many of these designs can be summarized using the (Figure 1) block diagram, which represents a crude superset diagram (not all blocks are in all designs). The diagram represents a dataflow processor, with application data entering the left side of the diagram through a FPID. The FPID permits more optimal shuffling of individual signal lines from the data stream to the “processing core”, represented in the center box of the diagram as a set of FPGA, memory, and first-in/first-out (FIFO) components. FIFOs provide for a level of separation between digital elements operating at different clocking / activity rates (like a sort of “temporal slip ring”). The configuration data for the different reconfigurable components is distributed through a separate configuration management processor (CMP) using an ad hoc local bus. In space systems, however, reconfigurable concepts suffer from the potential of upsets in both user and configuration memory due to the space radiation environment. Upsets in user memory can be compensated by design redundancy, but configuration memory upsets result in mutation of the form of the circuit itself, which is of great concern. Some

research in radiation-hard FPGAs continues, but the gap between hardened (0.03 M gates) and commercial FPGAs (8M gates) is staggering [20]. Rad-hard FPGAs may not be universally required, if brief outages (<1 second) can be tolerated in a digital function. In these circumstances, it may be possible to qualify the use of commercial FPGAs. To do this, it is necessary to show that FPGAs do not have destructive latch-up, can be shielded against total ionizing dose, and that any upset of configuration memory can be detected. Availability of dense, reconfigurable gate array technologies as previously emphasized are expected to play a central role in reconfigurable systems. FPIDs are like a special case of FPGAs in that they are interconnect-only programmable "switch-boxes" [52].

Reconfigurable analog elements— Digital-to-analog convertors (DACs) and analog-to-digital convertors (ADCs) provide an obvious bridge to the analog world, but all-analog techniques are additionally required to manipulate signals effectively. The traditional bilateral passive circuit elements (resistors, capacitors, inductors) are obviously crucial in any an analog circuit design. Integrated circuits can emulate variable resistors using conductance modulation approaches or from "brute-force" lithographic patterning of metal or polysilicon. Small capacitors can be readily formed. Inductors, larger capacitors, and high-precision resistors require non-monolithic techniques. Once again, MEMS technology offers possible solutions, including techniques for the construction of variable capacitors [26] and inductors [27]. Other ad hoc approaches can be employed, such as the construction of binary select networks, such as illustrated in Figure 2, which is an extension of a simple concept based on the combination of large, external capacitors combined with programmable switches. Combining MEMS switches with large, discrete components can create programmable bulk filtering, as suggested in Figure 3. It is also possible to combines concepts, such as shown in Figure 2, with FPID devices (exploiting the potential to normally digital switches for analog applications [53]) to create a "spare parts cabinet" that can be patched into circuits as required. Performing this function with transistor switches has the drawback of introducing a 50-150Ω series resistance for each switch, which can be reduced by using lower resistance (< 0.1Ω) versions of FPIDs based on MEMS relays, as suggested in [22]. Circuit elements can be conceivably formed from chalcogenide to form variable, programmable resistances, establishing an approach for extremely compact reconfigurable analog resistance networks.

Analog systems usually refer to multi-valued(>2 state) /continuum systems involving sampled-data and/or continuous-time measurement and manipulation of signals. There are many ways to introduce “knobs” into analog electronics. Not all concepts are new: examples from the past include a rich variety of circuits, such as programmable gain amplifiers, voltage-controlled oscillators, and tunable filters. However, as in the case of fault-tolerance, these tools are further extended through the concepts of reconfigurable systems. A reconfigurable systems version of such circuits exposes their settings / states to outside measurement and control, specifically to 0-1 decision processes (software control). Hence, it is possible to develop a great variety of analog building blocks with software-defineable interfaces, a collection of which is a reconfigurable analog system. Signal amplifiers, motor control, neural networks, and signal resolvers and converters are all examples of analog-domain systems that could be modified through the introduction of reconfigurable components system.

Analog is an extremely broad domain, which can be sub-divided in terms of signal level and signal frequency. Analog architectures, inspired by the digital FPGA, has been studied and a number of commercial devices have been introduced. Analog circuits, as a deliberate arrangement of elements, can conceivably be formed with programmable elements. Hence, a number of traditional commodity circuits can be built as reconfigurable circuits, such as amplifiers, oscillators, and filters. A number of researchers have been exploring field programmable analog arrays (FPAAs) [54,55,56] and field-programmable system-on-a-chip (FPSOC) [20]. [57] discusses the differences in programmable digital and analog, asserting that analog programmability is mostly parametric adjustments of more elaborate versions of fixed building blocks. Some approaches are clearly inspired by FPGAs. For example, [60] describes configurable analog block (CAB). Each CAB has a number of elements such as an op amp and switched capacitor arrays and a programmable input/output (I/O) structure. Complex analog functions (such as filters and closed-loop controller circuits) can be formed by combining multiple CABs. FPAA concepts have even been examined for use in space [34]. FPAA concepts are interesting, but have had relatively little success commercially. Among the cited problems are limited frequency range and reliance on external (usually non-reconfigurable) discrete components. Such concepts have even been studied for space, however, the performance of the commercial devices has been limited, and the range of applications narrow, due to the limited range of variations possible in analog components. However, the advent of improved reconfigurable components hybrid/MCM packaging of and improvement EDA concept makes possible the development of new, powerful classes of programmable analog architectures. With such blocks, more complex analog functions can be replaced by a compact module that can be reconfigured in system.

But these analog systems do not have the same expressive fluidity of their digital counterparts. The aforementioned attempts to make the analog equivalent of an FPGA have failed to achieve the level of acceptance that digital FPGAs currently enjoy. Part of the problem is in the much wider diversity of analog systems in frequency and dynamic range. It is not likely that a low-noise instrumentation amplifier could be reconfigured into a motor control driver, or that a thermocouple signal conditioner could be programmed to receive X-band radar signals. As such, it is necessary to consider the division of analog into the sub-domains: microwave, power, and (lumped-element domain) general instrumentation (for example). Hence, for this discussion "analog" targets the latter classification.

One attractive possibility for reconfigurable analog systems is to extend previous concepts through the addition of MEMS. At the most basic level, many continuous circuits are based on linear, bilateral passive components, such as resistors, capacitors, and inductors/transformers. If these components can be altered under software control, then the circuits based upon them are reconfigured.A number of MEMS concepts have already been proposed for tunable capacitors, inductors, and resistors [21].

MEMS switches can be useful in an obvious way to form programmable circuit networks. As a simple example, a linear array of resistors in the sequence of R, 2R, 4R, 8R, … can be tied together in parallel through MEMS switches as shown in (Figure 2). By opening / closing switches a discretely variable resistance element is formed, which can be viewed as a passive D/A convertor. With a large number of MEMS switches, combined with semiconductor-based switches and elements, it is possible to consider a potentially more universal / reconfigurable analog systems building block, along the lines of a "patch panel" system-on-a-chip. As in the analog computers of a more distant past, new configurable versions, supplemented with the rich base of reconfigurable circuits based on non-linear (active device) principles, stitch together disparate building blocks using patched connections, in this case formed by selectively switching on-die interconnections with MEMS and/or solid state switches, under external configuration control.

Since digital systems represent states with real voltages, and different digital systems can use incompatible voltages to represent these states, it is necessary to consider reconfigurable analog strategies to bridge systems together. The concept of *agile bi-level transformation* (ABLT) refers to the concept of using reconfigurable transceivers to support bridging digital systems based on incompatible signaling standards. Fixed versions of these transceivers are commonly available for connecting different legacy components together in modern design. In ABLT, each side of a bi-level transceiver can be programmed independently with distinct and (nearly) arbitrary voltage / current levels for logical "0" and "1".

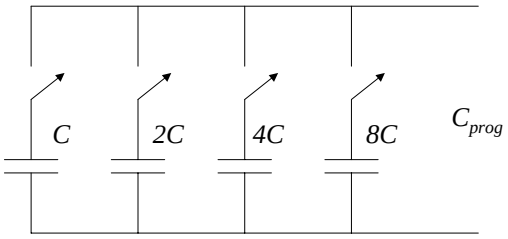


Figure 2. Simple, 4-bit programmable capacitor based on non-monolithic capacitors and programmable switches. The switches are implemented with transistor or MEMS-based elements.

Reconfigurable Microwave Elements— In the microwave regime, a wide range of approaches have been examined in addition to the previously discussed analog lumped elements that could be considered reconfigurable techniques, including tuning structures for transmission lines, reconfigurable antennas, and other reconfigurable filters and microwave circuits. A primary set of techniques for reconfiguring microwave circuit elements is based on the use of MEMS components: switches, resonators, and adjustable filters. Using circuits similar to Figure 2 with MEMS-based switches and replacing capacitors with delay lines for example, is a basic structure for a reconfigurable phase shifter. MEMS structures for tunable filters, resonators, and oscillators have been studied [35,36,37]. Chalcogenide materials may even have a role in the creation of reconfigurable waveguide and antenna structure [38]

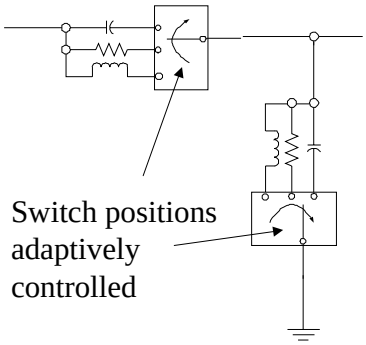


Figure 3. Example of how programmable switch configurations (in this case SP3T) can be blended with discrete components to make programmable filters.

Reconfigurable Power Elements—Power needs in terrestrial systems have become increasingly diverse. Complex systems typically support interfacing a variety of digital components that operate at different supply voltages. To combat this phenomena, a impressive variety of very efficient dc-dc convertors, low-drop out regulators have emerged and standards for managing power, such as the Advanced Computer and Power Interface (ACPI) [39] have been developed. ACPI is useful, because it exposes "knobs" for software control, permitting users to swap, for example, processors in a computer system without concern over its particular supply voltage requirements. Reconfigurable approaches, such as these, can in principle be applied to space systems to improve the ability to match new components with legacy systems, for example. These approaches can be further enhanced through the development of reconfigurable power components. An example of a potential component, illustrated in Figure 5 is the reconfigurable transformer, based on the combination of efficient planar magnetics and MEMS-based switches. Such components might be used to engineer a new class of "smart" power systems that can be agilely adjusted to meet dynamic changes in interface requirements while preserving power conversion efficiency. Power electronics, as a domain, is roughly delineated from other analog domains based on current and voltage levels considered extreme relative to minimum geometry devices on an integrated circuit. In “non-power” electronics, signals are often manipulated for the purposes of information management, whereas power electronics are more usually concerned with conveying an output representation to the “real world”, at levels ranging from milli-watts (this lower bound creeps downward over time) to mega-watts. Motor drive optimization and reducing noise generation are examples of power analog domain challenges. Normally, high-power systems are fixed due to the difficulty of achieving flexibility while maintaining high efficiency in power delivery. Reconfigurable systems approaches can be applied to make power analog functions more flexible and robustness with potentially minimal impacts to efficiency. An obvious method is based on the use of MEMS switches for component programmation. For example, bulk filtering can be made reconfigurable by implementing a superset of components in its pass and shunt networks, and selecting the desired ones under program control. This concept is illustrated in (Figure 5). The use of planar magnetics and MEMS switches make an interesting combination as a possible reconfigurable transformer, in which latching MEMS switches are opened or closed to realized different winding configurations. Such approaches can be used in programmable convertors and isolators.

Programmable Wiring Harnesses— As MEMS-based bistable latching switches (such as shown in Figure 4) emerge, they will permit the possibility of programmable wiring harnesses by directly embedding these elements into physical interconnection structures such as connectors, cables, printed wiring boards, and packages. Through the judicious arrangement of an adequate quantity of such switches, it is possible to establish routing networks within a space system flexible enough to dynamic connect power and signal connections between the various subsystems, making the satellite itself analogous to a large field programmable device.

Programmable Mechanical Elements— Multistate actuators, shape memory alloys, and other structures can be potentially harnessed in a reconfigurable systems to permit physical self-assembly, dis-assembly, and other actions as required. Programmable mechanisms can be based on arrangements of traditional actuators, although recent work has been reported on the use of distributed MEMS actuators with embedded control to make a mobile form of smart material, with programmable locomotion [41] characteristics. It is possible to envision the extension of such techniques to create programmable attachment concepts, which allows the programmable bonding of surfaces, programmable alignment structures, and other techniques in mechanical self-assembly under program control.

Programmable matter— Affecting the configurations of nanoscale structures to promote controlled self-assembly or the macroscopic expression of common physical or material properties are examples of programmable material concepts. Toffoli [42] identifies a potential control mechanism for programmable matter, based on performing highly dense computations in molecular-scale cellular automata substrates. The equivalent of "nano-software" is conveyed into the substrates and a computation process is triggered. The substrates could be used to convey the state of smart material systems in a distributed and finely granular fashion. Binding nano-activities to state transitions in such an automata system could establish a means to effect material properties at many point locations, achieving in effect a programmable material surface.

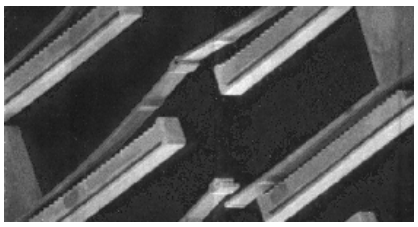


Figure 4. These thermally-actuated bistable MEMS relay based on LIGA fabrication demonstrate a high-performance contact structure for configuration applications. The contacts of two complete relays are shown, the lower programmed in an "OFF" state and the upper in an "ON" state [66].

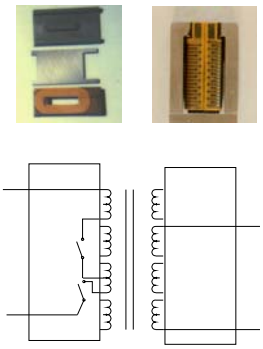


Figure 5 Advanced magnetic components (upper left) are combined to form flat planar inductors or transformers (upper right). The windings of inductors or transformers built in this manner are not fully connected, but can be finished with appropriate wire jumpers. One configuration possibility involves the use of MEMS switches (bottom) integrated into these structures to make software-controlled windings.

Reconfigurable Subsystems

Reconfigurable subsystems are based on assemblies of reconfigurable components. A few examples are briefly described here.

Reconfigurable computers—Computation can be thought of as a resource management problem. Computing tasks require a certain amount of time and space, and computer architectures establish within a fixed pool of resources an organization of how the pool can be focused to perform those tasks. The traditional von Neumann computers, as desktop computers are often described, involve substantial temporal reuse of a largely fixed set of resources. In some computers of this type, forming the logical AND operation on two bits or performing an integer multiply involve in a large sense the same resource pool, and the former operation is a poor utilization of those resources. Reconfigurable computing permits the more detailed management of underlying computing resources so as to maximize performance. Instead of sequencing the use in time of largely fixed resources, the entire set of resources can be organized in both space and time to perform more effectively certain types of operations. For this reason, it is not uncommon to see reconfigurable machines out-perform general-purpose processors by factors in excess of 400

or more on special types of computation [1]. At the same time, fully reconfigurable machines, particularly those based on FPGAs, are typical less effective for floating-point execution or processor emulation in general. Instead, it appears that FPGAs excel at certain types of processing, particular those involving fast-time control, pixel-oriented processing, and "bit-smashing". The line between an FPGA and processor is not crisp. FPGAs can be optimized architecturally for data-path processing or combined with processor cores, and general-purpose processors can be given more flexible resources in terms of processing sub-units, datapaths, and exposing the micro-architecture to users, as done in processors such as the wafer scale signal processor [43].

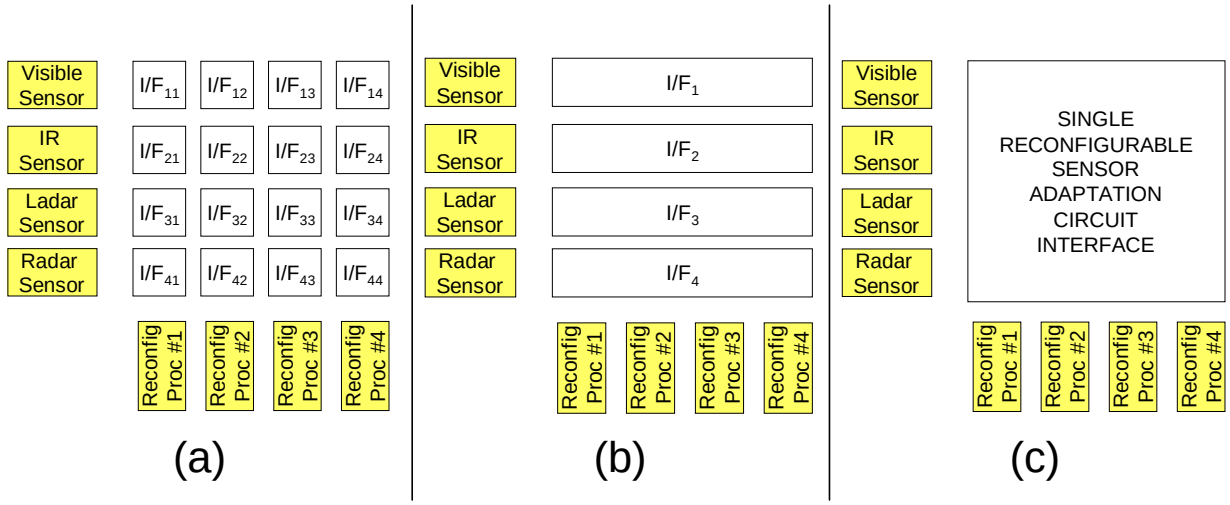


Figure 6. How reconfigurable analog sensor electronics reduce the proliferation of custom circuit types. In this simple example, the number of sensors to be interfaced and the number of reconfigurable processor types are both set equal (*N*=4). Without standardization of the digital interfaces of different reconfigurable processors, a worst case explosion of custom interfaces (N^2) must be constructed to connect any sensor to any processor (a). By standardizing the "morphable" reconfigurable processor interface, the number of custom interfaces is dropped, but *N* separate interfaces must be constructed (b). If reconfigurable techniques can be applied to the design of the sensor interface, it is in principle possible to construct a single programmable sensor adaptation circuit interface (c).

Software-defined radio (SDR)—As in the case of fixed-architecture computers, most radio-frequency (rf) transmitters and receivers are fixed in architecture. If digital processing is made fast enough, many of the standard rf operations can be performed digitally, provided the digitization is done early enough in the receiver (ADC) or late enough in the transmitter (DAC). The current physical limits in dynamic range and sensitivity of rf waveforms do not permit in general the notion of digitization at the antenna portal, which is the theoretically ideal objective of a software-defined radio, but many non-digital portions can be eliminated. Such an arrangement permits filtration, modulation, demodulation, and other common signal processing steps to be performed algorithmically. New protocols and modulation schemes can be implemented without removing or replacing components physically, and in fact a SDR can multiplex its internal configuration rapidly between multiple disparate rf sources and successively refine, alter, encrypt, or otherwise change its fundamental operations as a function of virtually any discernible parameter. It is even possible for a new configuration of the SDR to be transmitted to the SDR itself, allowing for novel possibilities in field usage [62].

Reconfigurable sensor processor—Concepts for open systems (as viewed by the Open Systems Task Force [45]) appear to rely on some backbone of well-defined legacy standards to achieve the desired overall goal of interchangeability. Concepts for reconfigurable systems can transcend this limitation[1], but most current approaches

[1] For example, a reconfigurable machine could implement RS-422, RS-485, I^2C, CAN, and a variety of other interfaces with software-only changes.

can only do so for digital signals, and even then only for digital signals that are voltage-compatible. Unfortunately, this limitation precludes the direct use of these concepts by many complex sensors and actuators, especially the more development versions of high-performance sensors. Interfacing most of these sensors to even a reconfigurable computer requires the physical construction of custom electronics. If a sensor is changed, then a new set of custom circuits must be developed. Even with reconfigurable computers, this creates an explosion of custom hardware. In general, for N different sensor types interfacing to N different reconfigurable computers, N^2 versions of custom hardware are required (Figure 6a), since no standards for reconfigurable computation interfaces exist. The situation is markedly improved if a standard for "interface morphware" can be defined. If a standardized back-end interface for the custom sensor front-end electronics hardware could be defined, then the number of custom electronics hardware interfaces drops from N^2 to N (Figure 6b). Since each of the N interfaces could be conceivably recessed into the sensor interface itself, this use of a standard morphware interface could eliminate the need for any additional custom hardware at this fairly high level of abstraction.

An alternative concept, which could reduce the number of interfaces from N to substantially (theoretically a single interface, as shown in Figure 6c) involves the notion of a reconfigurable sensor interface, referred to as a *sensor adaptation circuit*. Requirements for implementing such a sensor adaptation circuitry would involve defining digital and analog circuits capable of being reconfigured in system. The configuration of this sensor adaptation circuitry would be altered based on the type of sensor "plugged" into the circuit. Early concepts for such reconfigurable adaptation circuits may still need to be domain-specific, i.e., geared for a class of sensors. Using imaging focal plane arrays (FPAs), as an example, the generic types of required circuits are illustrated in Figure 7a. Such FPAs require a number of adjustable precision (static) bias voltages, a range of pulsed clock, sync, and mode control signals, and an interface to receive a number of parallel sampled-output streams, representing image data, as well as other digital or analog status information. The voltage levels for each set of signals may be completely different, with little standardization amongst FPAs, even for the same manufacturer. For bi-level signal conversion, a number of level shifters to and from the MSP are required, along with potentially the need for single-ended to differential conversion. For bias voltages needed by the sensor, a number of fixed bias voltage generators can be built from reference-driven operation amplifiers. These op amps serve two purposes: (1) tuning of particular voltages from fixed references, and (2) additional immunity from noise signals present on power supply lines. Analog-to-digital convertors (ADCs) are also employed to convert signal data from the sensors to digital bit vectors of 12-bit resolution or greater, which can be processed directly by a reconfigurable systems.

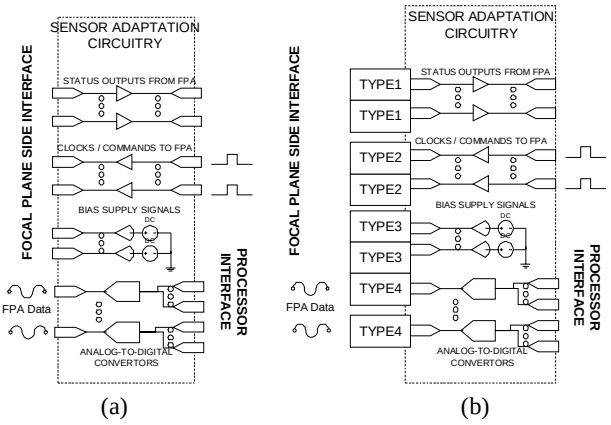


Figure 7. Typical FPA interface circuit elements. (a) Fixed version. (b) Reconfigurable.

The reduction in quantity of interface circuits come about when normally static interfaces in Figure 7a are extended through reconfigurable techniques, as suggested in Figure 7b. Four "types" of reconfigurable circuits are introduced, based on techniques previously described for reconfigurable analog elements. Requirements are outlined, but specific designs shall not be detailed here. Type 1 interfaces convert sensor digital discrete signals, which may be differential and/or non-standard voltage levels to formats compatible with a reconfigurable processor. Type 2 circuits implement the converse case, i.e., control signals originating from the reconfigurable processor to formats compatible with the sensor. Type 3 circuits are used to generate precision programmable dc voltages, for use in bias circuits. Finally, Type 4 circuits are used to perform signal conditioning and digitization of sampled analog waveforms. It is important to emphasize that analog signals as they are naturally generated are rarely in a form convenient for immediate conversion to digital format [58]. Type 4 circuits implement programmable gain, offset, and filtration functions necessary to make effective use of analog-to-digital convertors.

Reconfigurable control network—Reconfigurable concepts can be applied to avionics bus networks to enhance their flexibility, performance, and fault tolerance. Standard multi-drop busses, such as shown in Figure 8 have been used in aerospace systems, most notably in the MIL-STD-1553B serial (1Mbit/sec) bus.

Reconfigurable techniques can permit the extension of these concepts through the addition of even more copies of the primary bus, and the bus master can assign subsets of bus elements to connect to different copies of this primary bus. By itself, this concept is of limited utility, due to the single point failure of the bus master. This concept is illustrated in a-b, in which a failure in the bus master effectively shuts down the entire bus. Through the

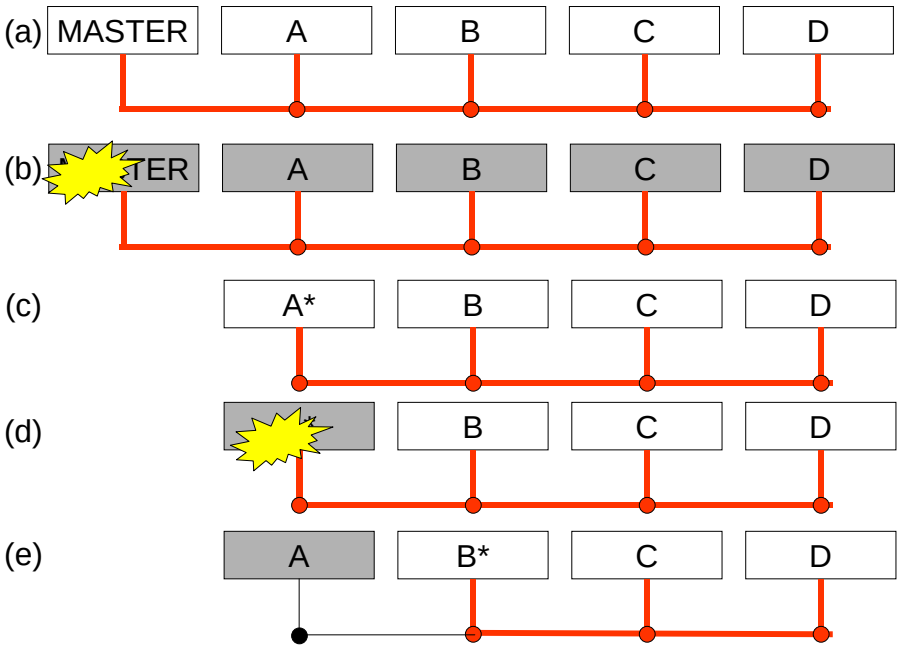


Figure 8. Overcoming the limitations of static mastering. In the static-master bus concept outlined the introduction of a fault in the master node (b) once again "takes out" the entire bus (dead elements shaded). In the dynamic-master concept (c), bus operations are distributed across elements, and a master node (denoted by the asterick) is "elected" through some heuristic. If this provisional master fails (d), a new master is elected through the same heuristic, enabling the remaining ensemble to continue operation (e) [47].

introduction of more advanced techniques, namely dynamic mastering, it is possible to use flexibility to increase robustness. The dynamic mastering approach can be viewed as a "headless" bus, in which a master is formed from the pool of elements attached in the network. Should this master fail, a new master can be assigned from one of the remaining nodes (Figure 10c-e). Of course, the dynamic mastering concept does not involve a special type of component, but rather is indicative of an approach based on reconfigurable systems concepts. Headless bus structures do not pre-suppose the existence of a particular master, but rather amortizes the functions of a master in some scheme across the pool of system "participants". More detail on a prototype implementation of such a scheme is discussed in [47] and a conceptual application is described in [48].

This concept is a special case of a much larger principle, one which de-emphasizes the supremacy placed currently on centralized schemes. Even for systems that involve a centralized computer, the concepts described here are important and applicable as a scheme to reduce the burden traditionally placed on a centralized process, by allowing an autonomous sub-network to manage and contain faults, with a looser coordination to the "real central computer" through any number of bridging interface concepts between the adaptive and non-adaptive networks.

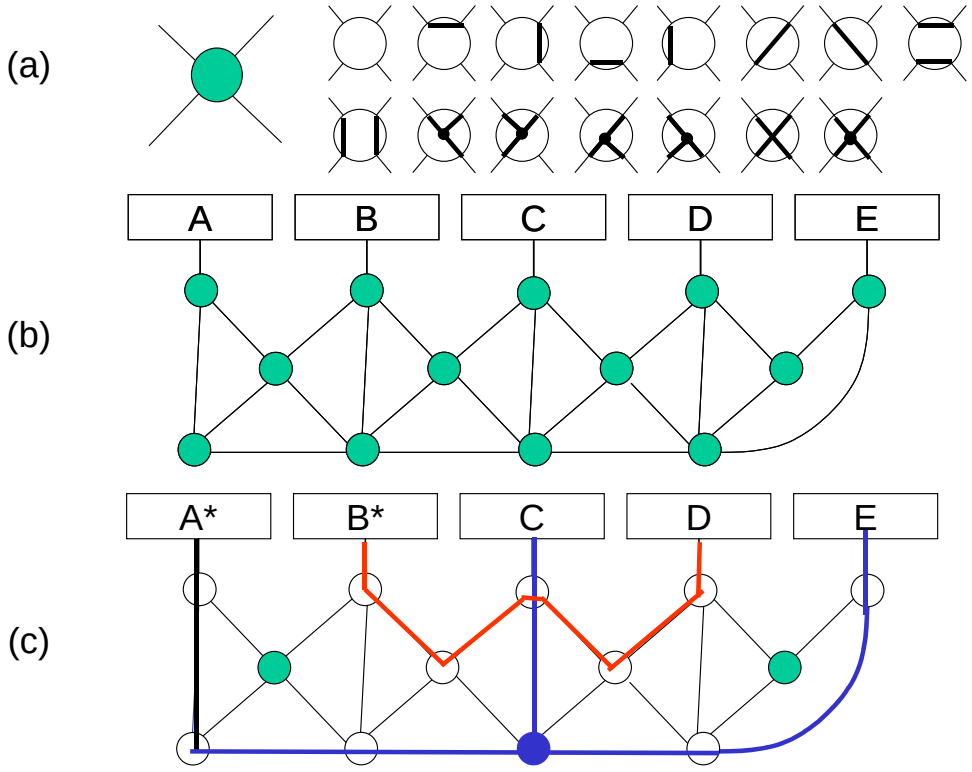


Figure 9. An illustration of interconnect switch fabric concepts applied to a reconfigurable avionics bus. The concept is based on the use of crossbar switches. An example four-port crossbar is shown in (a). The crossbar is non-blocking in that any conceivable combination of the inputs connecting to the crossbar shown can be dynamically set. With an interconnect based on crossbars, ad hoc network arrangements can be formed, as shown in (b). The fully programmable network is far more flexible than the redundant bus approach in that it is possible to dynamically partition the avionics into arbitrary sub-busses, each with dynamically-elected master nodes. In the example shown in (c), two sub-busses are formed, one with elements A-C-E (mastered by A*) and one with B-D (mastered by B*).

The concepts of multiple busses and dynamic mastering are valuable when applied in combination, but it is possible to do even better by making the bus structures more flexible. One way to achieve improved flexibility is through the introduction of a dynamically-configurable switch fabric. In this case, a programmable switch element such as a crossbar (Figure 9a) is used as the building block of a reconfigurable network (Figure 9 b-c). The enhanced flexibility resulting from this new programming wiring system, permits the simpler, redundant static-master bus concepts to be transformed into flexible, partitioned networks, in which failed nodes can be circumlocuted. Partitioning a single bus into several sub-busses also can enhance performance, since each sub-section can be operated independently and in parallel at the rate of a normal single or dual-redundant bus system.

Adaptive Hierarchy

The subject of what constitutes reconfigurability triggers debates, particularly when reconfigurable systems are touted as being superior to examples declared to be "non-reconfigurable". Another source of confusion in such discussions is the relationship between reconfigurable systems, self-organizing systems, and artificial intelligence. Part of the confusion may stem from a lack of clarity in distinguishing "knobs" from the way the "knobs" are adjusted. Figure 1 is an attempt to establish a taxonomy of the "knobs", but does not define a process for adjusting them. It is possible that the reconfigurable attributes of a system are managed externally, internally, or some blend of internal and external control. The *adaptive hierarchy* is an attempt to explain the degrees of programmability / reconfigurability and the source of control, ranging from fixed systems to autonomous systems capable of generating and refining behavioral goals. The hierarchy is viewed as "proper", i.e., each level assumes the existence of features from the preceding levels. For example, a self-configuring system (Level 2) assumes an underlying support of programmable "knobs" (Level 1).

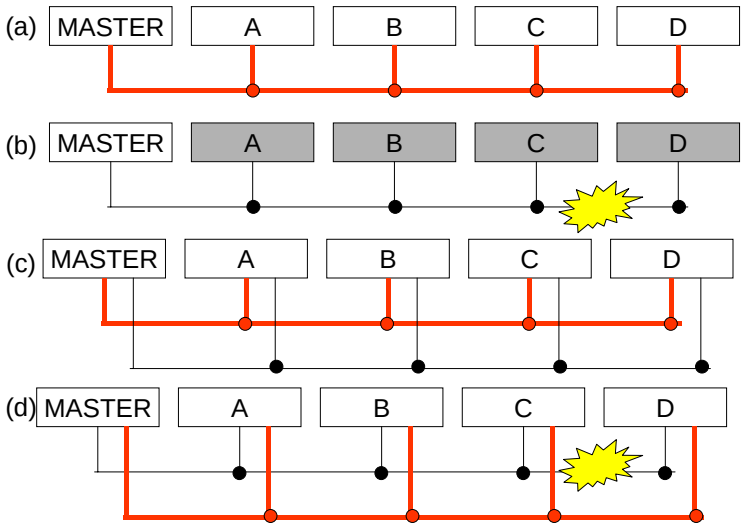


Figure 10. Fault-tolerance in avionics busses. In a static-master, multi-drop bus (a), a number of bus elements (A-D) are connected to a bus controlled by a common static master. The bus itself can be a serial bus, such as RS-485 (multi-drop version of the RS-422 bus), or it can be a parallel bus, such as VME. A bus fault (b) due to a failure on any bus element can potentially "take out" the entire bus, resulting in a catastrophic failure (dead elements shaded). To overcome the effect of a single fail, a second copy of the multi-drop bus is introduced (c). In this case, a bus fault (d) causes the failure of the primary bus, but the various elements can operate on the secondary bus.

To make the ideas of different adaptive levels more clear, simple, familiar examples and crude analogies will be employed. Even with these delineations, caveats apply. Obviously, both of the concepts of taxonomy and "adaptive-ness" do not necessarily relate to the degree of "infiltration" of a particular system by attributes that are software-defineable. It is conceivable that a reconfigurable could possess a single binary "knob" or billions of "(nearly) continuously-variable" ones. There is no obvious formula for declaring which is better in a given circumstance. One can imagine different "qualities" of reconfiguration, in ways perhaps obvious (i.e., the quantization or number of settings) as well as non-obvious (the number, type, and range of knobs needed to make a truly universal software radio). From this standpoint, no system is perfectly fixed, and no system is perfectly reconfigurable.

Fixed configuration – Level 0

Fixed systems are systems that are not designed to intrinsically support reconfiguration, or a system whose reconfigurability is sufficiently masked or inaccessible to make its classification otherwise difficult in a particular context. The classification is not always easy to make. For example, digital ASICs are sometimes considered fixed, whereas SRAM-based FPGAs are considered reconfigurable. However, within a very narrow domain, many ASICs, such as digital are quite programmable and flexible. In this sort of example, an ASIC will outperform an FPGA that is configured to mimic the ASIC, and for a number of parameters relative to its operation as a filter, the ASIC is programmable. However, the FPGA might be reconfigured to generate Bessel functions, and the ASIC cannot. So, the definition of "fixed" is itself context-dependent.

The usage context also plays a significant role in classifying a system as fixed. For example, most FPGAs are programmed once and used in an end system without the ability to support modifications. Conversely, a system design based on fixed components may be classified as reconfigurable, if provisions for changing performance and definitions of certain features are made.

Programmable configuration – Level 1

Programmable systems have "knobs" or features that are accessibly modified under electrical / software control. Within this level, a number of important characteristics can be considered:

- Discretization level (binary to continuous);
- Persistence (memory-less to non-volatile);
- Cycle-ability (one-time programmable or repetitively changeable)
- Dynamic-ness (the latency for changing features relative to operation)
- Partial reconfigurability (the ability to access selected subsets of "knobs" for change without disturbing other ones)

A particular group of settings in a reconfigurable system is referred to by a number of terms; configuration, programmation, personality, and bitstream are common examples. The development of these *programmations* involves the determination of "knob" settings to provide a desired functionality. The transmittal process of a programmation into a reconfigurable system is referred to as downloading, while the actual transferal of settings to all individual knobs is referred to as "configuring" or "installing".

Determining exactly how each knob should be set is in general computationally intense. In the case of digital configurations with even a single optimization criteria, such as minimum power consumption or maximum frequency, it is necessary to explore an exponential space defined by all possible knob settings. Most of these problems are considered to be non-deterministic polynomial-time complete (NPC), which is believed to require exponential solution time. This description of complexity applies, for example, to the determination of configuration bit settings in an FPGA, for example, for speed-optimized personalities. To make these computations tractable, even off-line, it is necessary to seek heuristic approaches, which generally compute a good, but sub-optimal solution to a given optimization problem in a much a much more manageable timeframe, such as minutes to hours (whereas the true optimization of an FPGA with only 500 different "knobs" would require a timeframe easily exceeding the entire recorded history of civilization by factors of billions).

As such, level 1 is considered to include the class of reconfigurable systems for which the personality or bitstream is defined "off-line". Level 1 reconfigurable systems operate using "canned" personalities, which are generally pre-computed externally and transferred into the system, usually well in advance of their installation. Also in the level 1 class of reconfigurable systems are the ones that support a "select-from-palette" reconfigurability. "Select-from-palette" reconfigurability involves storing of a number of pre-compiled bitstreams within an internal memory bank in way that permits rapid change-out. This concept can also be referred to as "context-switched" or "jukebox" reconfigurability, since like a music jukebox, such a reconfigurable system "plays" one of a finite number of "pre-recorded" (pre-recorded) bitstreams.

Even though, in principle, level 1 systems can be repetitively altered, the fluidity or agility of a particular scheme is heavily influenced by design of the configuration mechanisms. Two basic cases can be delineated. In *static reconfigurability*, configuration occurs only upon initialization, and the system must be taken out of operation to select a new personality. Static reconfigurability makes sense when reconfigurability requires significant latency (large FPGAs, for example, can require 100's of milli-seconds to configure) or when the penalty of reconfiguration is high relative to system benefits. In *dynamic reconfigurability*, the change-out can be done while the system is operational (on-line). The degree of "dynamicism" (how agilely the change-outs may occur) is a function of latency and the degree to which the following features are supported: (1) partial reconfiguration, (2) multiple-context configuration memory, and (3) in-system reconfigurability.

Another distinguishing attribute in reconfigurable systems is the mechanism or source driving the reconfiguration process. The possibilities include remotely triggered or locally (self) triggered. The latter case, which delineates the *self-configuring* system, can be established on the basis of scheduled (timeline-based) reconfiguration or event/data-driven reconfiguration. Timeline-based reconfiguration is practical in predictable mission profiles, such as target intercept, which involves a number of distinct phases that can serve as the boundaries for reconfiguration events. Missions of a more stochastic nature, on the hand, benefit from the ability to define reconfiguration cues based on circumstances arising during operation, such as a threshold exceedance in a data pattern.

Autonomous configuration – Level 2

Advances in the reconfigurable systems hierarchy above Level 1 are defined strictly as advances in computation or information technology, in other words, with "knob management." Level 2, for example, represents the ability for

systems not only to self-configure, but to *compute* partial or complete personalities on the basis of system alteration due to assembly or mission circumstance. For example, an unmanned air vehicle (UAV) is delineated from simpler radio-controlled systems by the sophistication of integrated sensor packages [26]. Future UAVs could conceivably compile the configuration of their avionics based on electronic descriptions embedded within sensors, activated upon plugging these sensors onto the UAVs.

One concept is referred to as *interface-driven* (or *assembly-driven*) reconfiguration. In this concept, an electronic data sheet (similar in principle to the transducer electronic data sheet defined in the IEEE smart sensor standards [27]) is embedded within a peripheral device or sensor. Upon connection, the data sheet is transferred to a host, which compiles the specification into bit patterns used to program the various reconfigurable elements of the system to accommodate the new addition. For simplicity, an memory-mapped host interface is shown, which permits the added elements to be viewed as regions of memory, for example, in a processor. Though the concept represents an interface-driven process as a one-way proposition (peripheral-to-host), it is not necessarily limited. In fact, it is possible to expand this concept to *both* sides of the interface, and in the limit, to all system interfaces. The advent of interface-driven reconfigurability can under such conditions be viewed as smart interface *auto-negotiation*, in which the various system elements iteratively exchange *possible* interfaces in search of a common ground for successful assembly. Clearly, this version of self-configuration is more elaborate still, since the specification itself can be altered to find a better fit. If processes such as these could be developed and perfected, they could dramatically reduce the time to construct complex systems and eliminate many forms of human error currently associated in the definition of interface control documents, which is often a source of mis-interpretation, mis-implementation, and integration delays in complex system development.

Of course, interface-driven reconfiguration is the level 2 equivalent of the static reconfigurability defined for level 1 systems. In general, once a system is assembled, the interface need not be recompiled. If the concept of electronic specifications is extended beyond assembly to more dynamic circumstances, then it is possible to consider missions as electronic specifications, data gathered from sensors as electronic specifications, etc. In other words, if the dynamic circumstances of operation can be translated into the equivalent of electronic documents, then it is conceptually simple to extend interface-driven reconfiguration into a dynamic form of self-configuration. As such, mission imperatives, specific data patterns from sensors, and recovery from faults can all become instruction sets to drive the dynamic recompilation and reconfiguration of on-board resources.

Fundamentally, however, the problems identified in level 1 systems remain, and most compilation processes can be NPC computation problems. It is necessary, therefore, to examine strategies to reduce compilation / installation time. Interface-driven compilation may not require real-time response, and may be less affected by these concerns, but the ability to reconfigure due to changes in operation force a more studied consideration of the heuristics and mechanisms for compilation (e.g., pseudo-linear algorithms, partial re-compilation, configuration caching).

Autonomous Goal Refinement – Level 3 and beyond

While complex and profound, the level 2 reconfigurable system can be built upon more or less standard hardware and software engineering principles. Even a Level 2 reconfigurable system can be viewed as collections of bitstreams, particularly when all reconfigurable resources are discretely represented. Considering higher levels of adaptiveness are possible, with increasing abstraction resembling in a very crude way the complexity hierarchies in computer science. In that field, questions of “what can be computed” (computability/decideability) and even “what can be recognized” can generate parallel questions regarding the limits of reconfigurability or (more basically) the limits of converge-ability. The association of these higher, adaptive levels to artificial intelligence concepts appears to be natural.

Level 3 in the reconfigurable systems hierarchy represents a set of concepts that, like an operating system, rely on the constructs of level 1 and level 2 reconfigurability as primitive “building blocks.” A level 3 reconfigurable system might appear intelligent, as it will be capable of optimizing its own reconfigurable resource pool in response to very high-level imperatives. Unlike the level 2 system, for which painstakingly detail specifications are needed to drive an autonomous compilation process, the level 3 system will be able to perform operations akin to program generation. Level 2 systems require scripts, and level 3 systems will form its own scripts based on its use environment. It may be that level 3 represents the meaningful intersection of “soft computing” disciplines (e.g.,

artificial intelligence, neural networks, fuzzy logic, genetic algorithms) with reconfigurable systems, though not the first[2].

Relation of the Adaptive Hierarchy to Modern Control Theory

Reconfigurable systems are "software-defineable" systems in the context of this chapter. Roughly, the concepts of observability (in the form of state monitoring) and controllability (in the form of adjustable knobs) apply. Methods for managing these "knobs" can be gleaned from control theory, but reconfigurable systems draw across a larger, more loosely defined set of concepts that may not always lend to the rigor of classical / modern control theory. For example, the correspondence between measurable and controllable quantities, does not necessarily provide the possibility of closure. The effects of adjusting some knobs may not result in changes that can be easily perceived or disentangled from observable quantities. The system "plant" model may not be adequately defineable and system objectives may not have a clear translation to optimization criteria.

Reconfigurable at an (Entire) System Level

Even as early industrial microcontrollers were transformed into a personal computer "appliance" through design, reconfigurable systems benefit from some unified design convention. Conversely, many reconfigurable components are simply "parts" without a framework. The common denominator of reconfigurable elements and the various rungs of the adaptive hierarchy is managing the configuration bitstreams. The CMF establishes a common structure for: (1) generating and distributing bitstream protocols for the various reconfigurable components in a complex system, and (2) managing their interconnectivity. As in the case of personal computers, a CMF can be evolved to accommodate progressively more sophisticated components and control concepts, and the core functionality can be refined into single chip form and distributed throughout a platform, localized to each cluster of reconfigurable elements. The concept of CMF evolved from previous work done by AFRL on the configuration management processors (CMP) used by the Malleable Signal Processor (MSP), as part of the previously described 3-D packaging framework. The CMPs allowed a number of different reconfigurable subsystems to be united using CMPs, which are networked together to permit the formation of reconfigurable systems of nearly arbitrary scale. CMPs provide a "wrapper" between the deeper details of a reconfigurable "client" and a large collection of these clients. CMPs comprise a standardizable "meta-interface," permitting the union of many desperate resources and the manipulation of their bitstreams. Any process of juggling, altering, compiling, loading, saving, and updating of the reconfigurable features of any client is reduced to a digital transaction involving one or more CMPs. The CMP extends the ideas of CMP to include a greater variety of clients and to provide eventual support for level 2 (and beyond) reconfigurable system concepts. The second function of a CMF, interconnectivity management, can be accommodated through the use of an adaptive manifold.

A simple proposed concept for a generic reconfigurable mono-system is proposed in Figure 11. This template unifies a number of previously described reconfigurable techniques, centered around a reconfigurable manifold, which can be used to flexibly define the terminal relationships between external connectors and a number of reconfigurable sub-units (digital, analog, microwave, power). A reconfigurable system of this form clearly extends the concepts of reconfigurable processing and adaptive computing to include other electronic domains. In this concept, a set of discrete passive components are linked with the adaptive manifold to extend the functionality of analog, power, and microwave sub-units in particular. Sockets are provided for interfacing the reconfigurable system to other traditional or reconfigurable electronics elements.

[2] Experiments had been done in the late 1990's on FPGAs whose configurations were successively refined through genetic programming to recognize different tones [8]

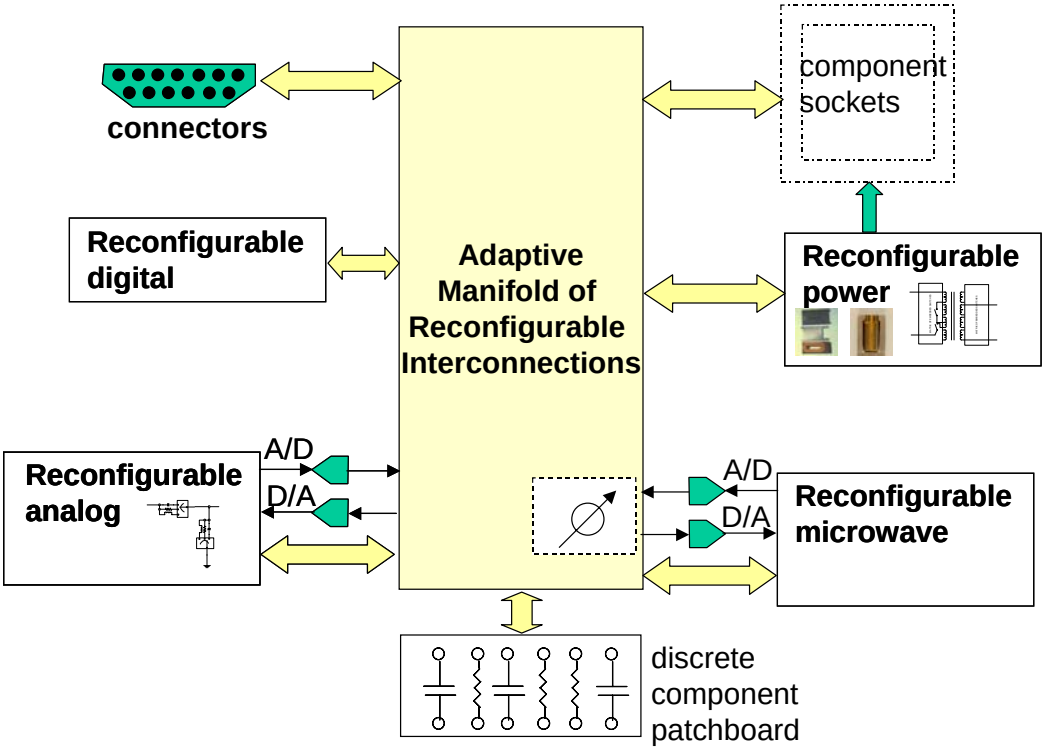


Figure 11. Depiction of a reconfigurable system, consisting of a reconfigurable wiring system and reconfigurable functional elements. This idealized view of a totally reconfigurable system is based on the unification of reconfigurable analog, digital, microwave, and an adaptive manifold of programmable interconnections. The adaptive manifold defines connections between the various peripheral blocks under software control. In this manner, even discrete parts and connectors, attached to the manifold, can be manipulated, pin-for-pin, in terms of their interconnection to other parts of the overall system.

To effectively bind together a set of reconfigurable resources, then it is necessary to merge together the control of the various "knobs" on each section. In the CMP approach, the reconfigurable portions of all elements in a system are exposed for manipulation by a single CMP through, for example, a boundary scan approach such as JTAG [65], which is commonly used for FPGA configuration. This methodology allows for a number of different reconfigurable systems to be united through a CMP-based control network, using concepts in Figure 9 to join the CMPs. The remainder of this subsection discusses a particular application in which many copies of the same reconfigurable mono-system are joined together.

Smart Satellite Panels and Cellular Spacecraft—The concept of a smart satellite panel is proposed partly to motivate a new approach for developing and configuring a wide variety of spacecraft systems very quickly using a few basic building blocks. Presently, spacecraft are engineered as individualized, custom systems, involving a number of complex, interactive subsystems, resulting in necessarily expensive and protracted development intervals. Very few spacecraft realize the economies in scale of more than a few dozen, and most are a production build of one. Even if the spacecraft structures could be built using LEGO®-like assembly concepts, the engineering efforts needed to add other custom spacecraft subsystems (power generation and generation, avionics, communications) would largely defeat the economies potentially realized. On the other hand, electronic systems-on-a-chip considered impossibly complex just ten years ago can be prototyped in a FPGA in just a few hours. Furthermore, the property of reconfigurability emboldens designers to try a variety of prototypes, each slightly modified from the last, repetitively as required on the same target hardware.

Under the premise that reconfigurability becomes more important in the future, it becomes necessary to consider both incremental and revolutionary design approaches to exploit the software definability implied by the term "reconfigurabilty". Of course, incremental approaches are easier, which are defined as enhancements to existing components, such as complex circuits (e.g., programmable filters), user interfaces (e.g., dashboards and displays), and subsystems (e.g, a telescope, receiver, or analytic instrument). Domain experts (component and subsystem designers) are in the best position to interpret progression toward "design for reconfigurability" in these respective domains. In the context of this discussion, revolutionary ideas are those without clear precedent or exemplar, i.e., they seem to come from "nowhere". They can be interesting, cross-cutting, even provocative, but they can also be risky, even heretical. In research, is important to study the revolutionary concepts in parallel with "safer" stepwise refinement of an existing base of concepts. Perhaps the greatest benefit in the study of these revolutionary concepts is not that they may directly find application in reality, but they may lead to important derivative concepts that can impact real applications. In part, this notion justifies the idea of undirected research in general.

It is with this prelude (and thinly veiled disclaimer) that the very revolutionary concepts of "protosats" are introduced. "Protosats" represent a concept for cellular construction that transcends many previous concepts of building blocks, because their construction is enabled by the ideas in reconfigurability that were discussed in the last section. Hence, structural diversity (as in LEGO$_{TM}$ blocks) only provides a visibly obvious foundation. Functionality must also be bound to this underlying structure. Though envisioned as a concept for future spacecraft, the idea of protosats could be applied to many different mobile platforms and fixed structures, such as buildings. Protosats provide a rich opportunity space for new ideas in packaging, MEMS, and architecture. The ideas described in this section motivate protosat simply as a creative possibility, and not necessarily define a rigorous roadmap for near-term development. Nevertheless, many of these ideas can be explored in research test beds and may provide valuable insights to other types of reconfigurable system components.

Multi-functional structures and Protosats

An earlier USAF program examined the concept of *multi-functional structures (MFS)* [29], which involved the enhancement of panels and spacecraft structures to include the electrical wiring functions normally implemented as custom cabling harnesses, active circuitry, and by implication, other possible functions, to include optoelectronics, fluidics, and provisions for thermal management. MFS was furthermore intended to support “plug-and-play”, or the LEGO$_{TM}$-like connection of structural elements together, with the implication that electrical conduits and other MFS features would be automatically managed in some overall scheme, perhaps automatically.

MFS was motivated by several basic ideas. First, the wiring harnesses of space systems are complex and expensive, and in large platforms, they are heavy. Since the cost of launch is sometimes rationalized on a per-pound basis, reducing mass could reduce launch cost. Furthermore, cabling harnesses are aesthetically cluttered, and they contribute to the outgassing problems in space systems by virtue of the surface area contribution of their polymeric insulating materials. Repair, upgrade, and refurbishment of space systems requires in many cases dealing with the complex manifold. By contrast, the MFS concept would minimize the size, weight, and complexity issues of complex manifolds by recessing the equivalent structures into structural panels. The concept would naturally reduce outgassing by reducing the contributing cross-sectional area, and complexity of pre-launch maintenance could by simplified.

The resulting hardware built in the early program, unfortunately, was a relatively meager and incremental step in the direction of panel-integrated functionality, which involved the adhesion of flexible circuitry to a panel, with a custom socket to accommodate a flat packaged hybrid microcircuit. Panel-to-panel interconnections were implemented not by separable connectors, but instead by pieces of flexible circuitry, resembling loops of tape that were attached with conductive adhesive. Repair, if needed, would require cutting apart these “tape loops” and reattaching them in one of several spare connection regions present on the flex circuit. The flexible circuitry in some cases required the attachment of supplemental wiring, especially for power distribution, since the current-handling of the flexible circuitry could not simultaneously support dense signal and robust power-handling requirements.

Though far short of achieving its original goals, the MFS concept provoked additional research in advanced packaging structural and functional architectures that could be merged with structural elements in spacecraft. Inspired by the LEGO$_{TM}$-like aspirations of the early MFS ideas and enhanced by other influences of advanced

packaging, VLSI, and MEMS, a new generation of concepts for smart spacecraft building blocks were conceived (Figure 12), aimed at not only overcoming the problems experience in the earlier MFS research, but progressing toward something like a universal framework for simple constructions based on smart primitive building blocks. These building blocks, referred to as protosats, would contain a set of "provisions" needed in a common denominator sense by all platforms.

Figure 12 represents an attempt to blend the divergent fields of reconfigurable systems with spacecraft design through the introduction of programmable tile architecture. The tile is based on the combination of a number of reconfigurable electronic and mechanical elements, using the Figure 11 approach of combining a number of reconfigurable resources. The tiles would be identical, but completely configurable using software approaches. As the name implies, a number of these tiles could be juxtaposed to form a system. Periodic flat assemblies could be used to form large, planar surfaces and three-dimensional shapes could be formed by using programmable, tile-joining fasteners capable of supporting a number of engagement angles.

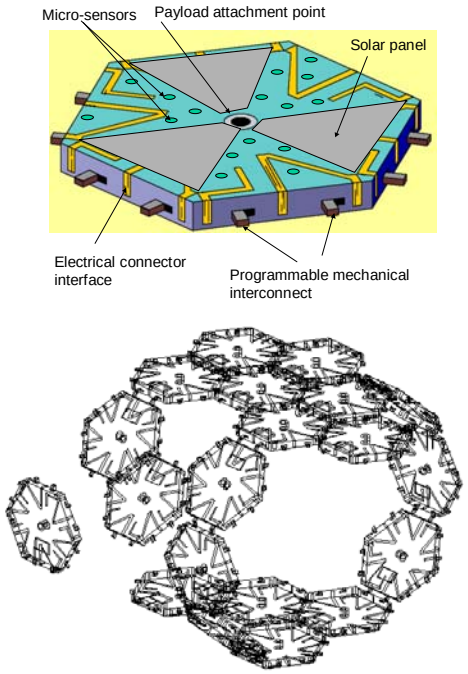


Figure 12. Smart panel tiles (upper picture), based on reconfigurable systems concepts, for use in the construction of cellular spacecraft (lower picture).

The protosat implements a single tile, based on combining a number of reconfigurable electronic and mechanical elements in a common-denominator structure. A proposed list of protosat features is as follows:

- embedded configuration management processor;
- a number of embedded sensors to measure health and status, such as thermometers and dosimeters;
- a reconfigurable power generation, management, and distribution subsystem;
- a reconfigurable wiring manifold;
- programmable mechanical contact systems and tile-to-tile interlocks with high-density, auto-aligning connectors to transfer interconnections between tiles;
- freespace optical transceivers for wireless communication between tiles;
- A primary payload mounting site

Other optional elements could include:

- Integrated propulsion system;
- Distributed reconfigurable RF antenna and supporting electronics;
- Generic built-in instrumentation resources and networkable fixed and/or reconfigurable processing.

The protosat tiles are designed with a number of edge facets, designed to attach to other protosats. The tiles would be identical, or at least drawn from a small variety of compatible tile designs, but completely configurable using software approaches. Complex space systems structures could be based on the juxtaposition of a number of these tiles to form a collective referred to as a *macro-sat*. Periodic flat assemblies could be used to form large, planar surfaces and three-dimensional shapes could be formed by using programmable, tile-joining fasteners capable of supporting a number of engagement angles.

Protosats, as building blocks, are the basis for more complex structures ("macrosats") and borrow concepts from VLSI and reconfigurable technologies. In the newer conceptual framework of universal construction approaches based on protosats, it is necessary to define some basic terms:

Cell Library. In analogy to VLSI, a "protosat" cell library consists of a set of resource definitions corresponding to the various available building blocks in the "protosat universe". Just as in VLSI design, where prebuilt cells can be chosen to include any floorplan layout based on a function, speed, size, shape, and drive current, is possible to analogously define primitive and non-primitive cells based on available protosat elements.

Configuration bitstream. The binary specification of set of protosats (and their resources) and their arrangements for a particular macrosat.

Macrosat. The skeleton of a complete spacecraft based on the assembly of primitive (protosat) building blocks. The structure of a macrosat can be represented as a unique configuration bitstream specification. The macrosat is referred to as a "skeleton" and not a complete spacecraft, because its definition does not include a number of the auxiliary (parasitic) attachments needed to customize the spacecraft for particular mission, especially the primary payload.

Modular attachment point. A region of a protosat allocated for the attachment of optional "third party" expansion functionality, such as customized processing.

Parasitic attachment point. A region of a protosat allocated for the attachment of load bearing structures. The distribution of parasitic attachment points is designed to correspond to an integer multiple of some basic feature length, permitting the standard design of parasitic "third party" payload, propulsion, and power generating facilities not provided in the standard complement of elements drawn from the protosat "cell library".

Protosat. A fundamental smart structure, from which more complex structures can be built. In addition to its structural "viability", the protosat is equipped with electrical, processing, and communication subsystems, as well as smart connector (mechanical and electrical) interfaces and parasitic structural and modular attachment points. In the current state a conceptual development, protosats may also eventually include solar power and simple propulsion resources. In an attempt to at least to confine rampant terminological abuse, it is not analogously correct to refer to the protosat as "atomic", in recognition of the possible existence of simpler, less complete building blocks (protosatoids) that may be necessary to form arbitrarily complex structures.

Protosatoids. A structure reduced in complexity as compared to a protosat, from which more complex structures can be built. Protosatoids are consistent with a cellular paradigm, but particular variants like one or more of the subsystems present in the base definition of a protosat.

Resources. Primarily refers to one or more of the pools of reconfigurable assets within a protosat. FPGAs, for example, can be defined as having logic, memory, and interconnection resources, each of which can be largely defined by a user/designer. In protosats, power distribution, signal wiring, microwave pathways, processing functionality can also be viewed as resources whose utilization can be defined by a user through automated design processes.

Relationship of protosats to nature and computer science

Protosats represent not only an abstraction of reconfigurable systems, but also demonstrate correspondences to self-assembly / organization in nature and computer science.

Viruses

Viral organisms comprise the simplest biological systems, and provide a number of interesting correspondences to simple self-replicating concepts in computer science, such as the computer virus. Some of the simplest viruses have icosahedral symmetry. One example, the cowpea mosaic virus (CPMV), is characterized completely as having a very simple biological "bitstream" (RNA sequence) consisting of 3000-5000 base pairs. The virus structure consists of a capsid or outer shell, which wraps around a payload consisting of a folded RNA strand that carries the informational sequence used in replication. Simple viruses are recognized as supporting at least two levels of self-assembly. The first level of self-assembly and the most intriguing is the sequence of actions which lead to the construction of another copy of a virus. Production of viruses requires the infection of a suitable host, since virus reproduction relies on "hijacking" the RNA production facilities of another organism. The payload responsible for this "hijacking" is contained within the viral capsid and is released when the capsid drifts into the appropriate region

of the host organism and dissolves, freeing the viral payload. The host's tendencies to produce its own RNA is then overwhelmed by the virus payload, which instead causes the host to produce large quantities of the primitive components that make up the virus. Those components include a number of primitive molecular units that are used to form the capsid and an RNA strand. The RNA strand serves both as a payload for the virus and as a trigger mechanism that initiates a self-assembly sequence. Once triggered, the primitive molecular units spontaneously self-organize to form the capsid around the RNA strand, presumably inducing its folding for compact storage within (on rare occasions, a capsid is formed without enveloping this payload, yielding an impotent virus by definition). Even though these simple biological self organization/self-assembly mechanisms are not completely understood, they perhaps represent a more suitable basis for mechanisms engineered for artificial self-assembly than the vastly more complex processes exploited by higher organisms. Could there be an analogy between the icosahedral capsid of the CPMV and a macrosat, both of which are formed by self-assembly processes?

Cellular Automata

Cellular automata (CA) are man-made arithmetic structures, which have been extensively studied as the basis of many natural phenomena. In their simplest form, cellular automata can be viewed as discrete credit compute points in one or more dimensions. Dish point represents a value, usually discrete, which is updated discrete points in time through estate transition matrix. The state transition matrix is defined slowly in terms of a certain local neighborhood and can be represented as an ordinary truth table. CA are interesting because, as a collection of extremely primitive building blocks, they can demonstrate unexpected complex behavior.

CA were proposed initially by von Neumann and Ulam in the 1950's, inspired in part by the search for self-replicating automata. Through many subsequent contributions (more recently catalyzed by Wolfram's investigations of their statistical mechanics in the and 1980's [30]), connections have been made between CA and physical phenomena, ranging from earthquakes, seashell and snowflake patterns, and even the origins of randomness in nature. The notion in exploiting CA for self-organization had been identified (for example) by Wolfram [30], who observed that particular global rules exhibit an evolutionary behavior leading to an equilibrium state structure that appears to be independent of initial boundary conditions. In other words, if somehow self-assembly primitives could somehow be made equivalent to those CA structures, then it might be possible to engineer deliberate forms of self-organization in artificial systems.

Reconfigurable Cellular Arrays

While viruses and cellular automata demonstrate self-organization based on simple, periodic structures, other work has examined the possibility of introducing specialization within such collectives. Reconfigurable cellular arrays (RCAs) are defined as simple periodic assemblies that support the property of site-specific behavioral specialization. While CA exhibit behaviors based on collectives of cells with identical behavior, RCAs exhibit behaviors based on collectives with similar spatio-temporal restrictions, but the ability to customize each state transition matrix. RCAs in a purely mathematical sense are not really more powerful than CA, but in a practical sense are more versatile, since they can produce equivalent functionalities more densely and demonstrate defect tolerance (i.e., through site specialization it is possible to map around defective sites / tiles). As such, RCAs are being studied as an architectural approach for molecular electronic systems, where it is necessary to create structures with the greatest expressive capacities using the simplest possible building blocks [31].

Design of satellites and structures based on Protosats

Protosat design and integration would directly benefit from tools similar to those used in FPGA design. The eventual hope is the emergence of a rapid prototyping scheme for spacecraft, in which a multi-year process could be considerably reduced. In principle, at least the skeleton of a spacecraft could be prototyped in hours, assuming adequate inventories of pre-built protosat tiles. The design / assembly of these protosats into macrosats and eventually complex space systems might adopt a sequence similar to the following:

1. Design spacecraft primary skeletal structure using computer-aided design (CAD) tools. The macrosat structure can be readily captured compactly as a graph structure, with nodes representing protosats and protosatoids, and edges representing the connections between them. As a graph object, the macrosat structure might be readily analyzed and further optimized through a number of consistency and viability checks, in other words, design rule checks. Structural simulations can also be performed, perhaps as an iterative process, to preclude mechanically inadequate designs.

2. A set of global subsystem resource maps are provisionally compiled to establish unified: (1) power generation, management, and distribution; (2) distributed health and status network; (3) guidance, navigation, and control (attitude determination); (3) command and data handling.
3. Modular and parasitic (payload) payload attachment points are defined throughout the skeletal (macrosat) structure. Parasitic attachments will typically represent resource sinks and structural deviations, whereas modular attachments can serve to replenish / balance such drains. As these are added, the subsystem resource maps are dynamically updated, and any deficiencies are highlighted as updates to the rule checking procedures. This part of the design flow provides an opportunity for enhancing the resource pools with redundancy for design margin.
4. More comprehensive compilations are performed of the various subsystem resource pools (e.g., wiring for power, control, and data between payload elements) using reconfigurable resource pool form by the ensemble of tiles, resulting in a compact bitstream that represents entire spacecraft;
5. Draw required quantities of tile from inventory, assemble as required, and "inject" bitstream into final structure, which finalizes the spacecraft configuration
6. Perform macrosat construction.
7. Add parasitic elements to macrosat.
8. Perform additional reconfigurations as required to finalize integration. The need for such recompilation / reconfiguration may be driven by residual errors in third party elements.

The basic macrosat construction is envisioned as a ground assembly procedure. It is possible to envision a number of compelling variations to this assembly approach. For example, the “smart tiles" could be stockpiled in orbit and released to initiate an *in-orbit* construction of a spacecraft. The tiles could use a self-contained auxiliary propulsion system to self-assemble without manual assistance. The final spacecraft structure could be realized through a beneficent analog of the concept of a virus, in which a blank tile is "infected" by transferring the bitstream of the entire spacecraft into its local configuration memory. As it encounters other tiles (communicating through a wireless linking approach), they are likewise "infected" and join to preceding tiles, resulting in an incremental, piecewise assembly of a complex system. Reconfiguration cycles are also performed incrementally, eventually approaching a finalized configuration. In this sense the viral analogy might also be interpreted as a genetic analogy, as the complete description of a spacecraft might reside in every tile.

These cellular spacecraft have other implications to economies of spacecraft servicing and asset protection. For example, it is possible to "shed" defective tiles and replenishing with new tiles (akin to biological regeneration). The new tiles could be held on-board as spares or re-supplied from some constellation of on-orbit repositories. Instead of de-orbiting a no-longer needed cellular spacecraft, its components can be dis-assembled and re-used to supply other spacecraft construction efforts in-orbit. It is conceivable that in some cases, a cellular spacecraft that might be under attack could "dissociate" and temporarily disperse its many components until the threat subsides (under the premise that its many pieces would be harder to destroy than an integral spacecraft). The cellular "paradigm" for spacecraft could provide many benefits to on-orbit servicing, space logistics, and other space population /colonization concepts.

Progress Toward ProtoSat Support Concepts

While a comprehensive proof of “universality” or “correctness” is not presently possible for protosats, it is straightforward to demonstrate several elements of the basic approach. For example, it is possible to form a basic “smart-assembly” mechanism for structures based on the genetic / cellular notions previously suggested. The concept of an adaptive manifold, a primary mechanism for implementing reconfigurable wiring resources, can also be implemented in the near-term based on the emergence of MEMS latching micro-relays.

A small-scale internal program at AFRL has created a simple proof-of-principle demonstration based on small printed wiring boards (PWBs), which represent protosats. In this case, the “protosats” are small, rectangular assemblies with essentially no structural load-bearing capability. Connectors exist on each each in three orientations (up, down, and straight out), and each board contains a small microcontroller with non-volatile memory. Several dozen copies of this design have been constructed, and a macrosat is initiated by programming any single tile with an initial bitstream representing the structural specification of the entire macrosat. Additional “protosats” are then randomly plugged to the initial (seed) board by humans. A series of light-emitting diodes (LEDs) present on each edge emit a green color only when boards are plugged in the correct orientation. Incorrect guesses, which do not

produce illumination, serve as a indication to unplug the given board and attempt an alternate orientation. The correct assembly of subsequent boards forms a sort of expansion front which grows until the final desired macrosat structure is realized.. As in the genetic analogy, the bitstream (structural specification of the macrosat) is completely contained by each tile upon its correct connection to the expansion front.

Adaptive manifolds, as described in the previous section, represent at least one other concept that is accessible with recently available technologies. Each protosat could in principle be embedded with both fixed and configurable wiring resources, which are connected throughout the protosat and various interface points. As in the case of the structural specification, the manifold can be represented as a graph. Cascaded tiles could be represented by joining the subgraphs of each component protosat. Preliminary work towards the adaptive manifold is described in [22].

Summary of MEMS and MEMS packaging Opportunities in Protosats

Protosats provide a number of clear research opportunities in MEMS and packaging:

- Integrated sensor clusters (e.g., pressure, chemical, thermal, vibration, radiation) for built-in health and status;
- "Smart Velcro" for structural attachment [32];
- Precision self-aligning electrical connectors (described in [33]);
- MEMS latching switches for analog, microwave, and power signal distribution using the adaptive manifold, preferably co-integrated into the wiring harness assembly;
- Determining optimum population strategies for such switches to minimize the number required in a given protosat;
- Low-cost inertial reference sensors for limited ability to support intrinsic attitude determination;
- RF MEMS for intrinsic communication;
- Optical MEMS as a possible means to communicate to physically separated tiles in close proximity

Summary

This chapter has provided one view of the potential of reconfigurable technologies to revolutionize the design, development, and use of satellites after they have been deployed. While at times the discussion has taken on a dimension that appears like science fiction (even in "2001"), each of the ideas are based on fundamentally sound premises.

The benefits of reconfigurable systems to reduce time in gaining space access from initial concept have been emphasized, along with the asserted expectation that flexibility is also an asset to systems that have already been deployed. If a spacecraft represents a collection of resources, and it is difficult to service these resources, then it seems reasonable to conclude that the ability to re-align, tune, or otherwise adjust these resources could extend the useful life of spaceborne assets.

A reconfigurable taxonomy was developed, along with a number of examples to reinforce how reconfigurable concepts could be applied to a wide range of components and subsystems. A roadmap of reconfigurable approaches might converge in the eventuality of a totally reconfigurable system, and a suggested generic representation of such a system was provided. As a final "reaching" extension of the whole set of ideas, the synthesis of a cellular spacecraft was presented in which a number of the elemental strategies for reconfigurable systems could be combined to form a fundamental different approach to constructing spacecraft.

While the practical use of reconfigurable systems technology may not take on the proportions of the Figure 12 vision, it is clear that reconfigurable technologies, at least in the form of reconfigurable gate arrays, are very popular terrestrially. The lack of suitably space-qualified forms has not stopped researchers around the world from trying to use the components. This may in part be due to the recognition of the power of flexibility in systems design and attempts to bring it into space. This paper has largely expanded on that theme with a set of concepts that may make systems of the 21st century live up to their 20th century expectations.

3.0 Reconfigurable Wiring Harnesses

Reconfigurable interconnections are fundamental to any field programmable architecture. They can be used, agilely and repetitively, to define complex, seemingly arbitrary patterns of linkage between each terminal in the thousands of building blocks contained within a contemporary FPGA [51]. This flexibility has given rise to the very realistic illusion that a designer can create new designs, in some cases containing millions of gates, and implement them directly in a piece of pre-fabricated silicon by simply shifting a stream of binary digits into a configuration port. Without this flexibility of interconnections, even the most powerful FPGAs are reduced to a loose, inelegant collection of memory cells.

At the scale of multiple integrated circuits, the need for the flexibility of interconnections in digital systems has given rise to a new type of programmable device, referred to as a FPID [52], in which the network relationships between the package terminal pins can be manipulated under program control. These devices represent one extreme in field programmable systems in that they have no user logic or memory elements, but represent silicon that is dedicated strictly to programmable interconnection routing. FPIDs offer a powerful new tool in systems at a complexity scale beyond a single-integrated circuit which can benefit from having a pool of several hundred signals whose pathway relationships can be rapidly re-defined as desired.

The obvious successes of FPGA and FPID devices demonstrate the importance of reconfigurable interconnections in complex systems. Unfortunately, FPGA/FPID devices are limited to applications involving CMOS-compatible signaling levels (implied in this is the switching and switched voltage levels are comparable in value and reference to a common circuit). Furthermore, while it is possible in a very limited way to apply FPID devices to analog [53] applications (limited frequency and signal excursion range), the most powerful tools for the construction of reconfigurable systems are, for the most part, limit to use in CMOS digital applications. It is reasonable to believe, however, that real world systems, which in addition to digital contain analog, microwave, and power waveforms, would benefit in the same way digital-only systems have from a reconfigurable interconnection approach. Unfortunately, conventional semiconductor approaches do not lend themselves as a medium for switches to handle signals with the levels of isolation, frequency performance, and excursion ranges. In FPGA/FPID devices, a programmable interconnect is essentially a software-controlled wire, based on a MOSFET switch/switch pair and actual physical interconnect. This function is achievable with a very classic "legacy" technology, namely the electromechanical relay. Indeed, traditional relays still find application in many contemporary systems, in roles where CMOS-based switches cannot provide adequate performance.

Is it possible to somehow apply FPGA/FPID concepts to traditional latching relays? In principle, ofcourse, it is possible to emulate FPID structures, but this is not practical due to the bulk and reliability of mechanical relays. Recent developments in MEMS technology have led (among other things) to the advent of new classes of micro-relays in both non-latching (momentary contact) and latching configurations. If relays of adequate quality could be miniaturized substantially, then it is conceivable that at least rudimentary FPIDs could be constructed in a chip-sized format in which signals from a variety of electronic domains could be intermixed and arbitrarily combined. Even at densities far below than that of contemporary digital microelectronics, it is likely that a family of general purpose FPIDs could be leverage into systems design to provide a new level of flexibility in systems. As will be described, the adaptive manifold is something more than an FPID, but in fact is a programmable wiring system, in which relay-based FPIDs are used as components.

It is upon this premise that the present paper is based, organized as follows. First, a motivational basis is provided for adaptive manifolds. Next, the basic concepts of an adaptive manifold are defined. Then, the requirements of relays for these manifolds are discussed and the state-of-the-art in latching MEMS micro-relays are reviewed. The subsequent section discusses some practical problems to be addressed in the development of adaptive manifold. In this section, the limits of homogeneous relays are discussed, leading to a hybrid system in which relay and silicon-based FPIDs of different types may be intermingled. Finally, the design of a rudimentary relay-based FPID is reviewed, based on a two-phase development involving at first conventional and then later MEMS-based relays for the switch fabric.

Reconfigurable Systems

The monochromatic Hollywood cinema depictions of scurrying human telephone operators plugging cables into patch panels represented perhaps the oldest example of an adaptive manifold. More recently came the patch panels of analog computers, still for the most part driven by humans. Vacuum tubes, transistors, and integrated circuits

rendered them all obsolete for the most part through electronic switching and high-performance digital computation. In many electronic domains, it is possible to convert a signal to or from a digital representation, which is easily manipulated and transformed by a staggering array of circuit and algorithmic techniques, often in real-time. In the cases where this is not possible, a far more limited range of options exists. It is not generally possible to change the dielectric properties of matter for example by setting a bit in a memory cell. Connectors usually do not self-mate or un-mate. Power supplies and crystal oscillator frequencies are for the most part oblivious to digital transaction, at least until very recently. Clearly, it is desirable to render as much of the real world accessible to digital influence as possible, since from one perspective, the digital domain is a sort of transformed existence, where actions are reduced to 0-1 programming decisions, accessible to algorithms and thought exercises. When a hardware engineer declares that a problem is "only software", it is not so much to trivialize software development as it is to acknowledge that a problem can be exposed to a much larger range of options, none of which require the cutting and soldering of parts and wires.

Before the advent of FPGAs, the "thought exercises" of software were confined to the manipulation of instruction sequences and registers in a stored-program computer. FPGAs dramatically opened up the space of possibilities by exposing directly the logic, interconnect, and memory resources of digital electronics completely to the whim of design software. Even though today's electronic design automation is far better able to insure first-pass success in the design of fixed integrated circuits through modeling and simulation, the adoption of reconfigurable circuitry continues to expand. It is clear that flexibility is important in digital design, and the need to resort to a dedicated ASIC development is diminished by continuing improvements in the state-of-the-art in FPGA technology.

Of course, FPGAs as they are traditionally discussed are digital systems. Analog FPGAs have also been attempted [54,55,56]. In some devices, analog programmability is limited mostly to parametric adjustments of more elaborate versions of fixed building blocks [57]. More modern concepts implement improved modular functionality [58], in the form of mixed-signal systems-on-a-chip (SOCs), some of which mix and interconnect digital and analog functions [59]. Some of the more recent concepts allow programmable filters (up to fifth-order, based on a library of over 8000 configurations) or complete program proportional-integral-derivative (PID) controllers[60]. Some approaches are based on coarse-grain [60] and fine-grain [61] programmable analog element structures, to implement a variety of functions, such as op amps, comparators, switched capacitor arrays, offset removal circuits, rectifiers, gain stages, filters, oscillators, equalizers, and pulse-width modulation circuits. One approach for programmable analog involves a more mathematical, function-oriented approach, reminiscent of the look-up tables (LUTs), which compute Boolean functions in digital FPGAs. In this case, a matrix arrangement of analog building blocks are employed, each of which can be programmed from a set of the following functions: open, short, add, negate, log, antilog, rectifier, differentiation, integration, amplification, and attenuation [57].

The analog versions of FPGAs are limited in flexibility and performance. The performance gap is greater than those in digital FPGAs when compared to custom logic. Flexibility remains an important characteristic, and examples of post-design reconfiguration have been discussed [60]. However, in most cases, it is necessary to resort to a variety of external components, which limits prospects for post-assembly configuration [56,61].

Approaches in reconfigurable RF/microwave electronic and antenna elements also exist. Some techniques are distinctly digital, but at much higher frequencies than typical digital applications, such as software radio [62] and direct digital synthesis schemes. Reconfigurable antenna approach involves long-standing approaches, such as ferrite-based electronically steerable arrays, along with newer versions based on MEMS phase shifters. A number of MEMS resonators, filters, and other tunable structures have been described for RF applications.

It is for want of flexibility in all aspects of electronic systems that provides a central motivation for the adaptive manifold. Adaptive manifolds refer to the combination of bistable, MEMS-based relays and traditional interconnection structures: packages, connectors, multi-chip modules, printed wiring boards, and cable harnesses. The characteristics of the manifolds can be defined at will under software control. Hence, a connector or circuit board could be in effect re-wired within a system, without the need to dismantle or even (in some cases) remove power from a system. Adaptive manifolds could be used to recover failed systems in the field, change functionality in system after deployment (extend or enable new missions), or simply allow for a vastly greater flexibility in the design, use, and operation of systems.

Space systems represent a special case in that they are not generally serviceable. The use of adaptive manifolds could close the gap between fixed and serviceable systems. In some cases, space assets could be remotely or autonomously reconfigured and refocused to adapt to changing mission needs in a way far more flexible than permitted through software only changes inside of an ordinary computer. For space systems, this would directly result in the reduction in the size and mass of spacecraft systems by maximizing the utilization of the individual components themselves.

The adaptive manifold is an enabling technology in the sense that it will contribute to the development of miniaturized, intelligent, self-repairing and self-programming sensor systems.

Architectures For Reprogrammable Wiring Systems

One concept for a fully reprogrammed system is shown in Figure 11. The adaptive manifold is central this idealized view and can be thought of as a binding agent for a number of otherwise fixed or reconfigurable component technologies. By "laundering" connectors, external and internal, through an adaptive manifold, it is possible to alter the wiring order under program control. Reconfigurable digital, analog, power, and microwave elements can be attached to each other in a way that is also reconfigurable. To overcome the limitations of some of the field programmable analog concepts, a "spare part bin" can also be attached to the manifold, which can be used to manipulate the terminal relationships of various large capacitors, inductors, and other circuit elements to provide for reconfigurable arrangements of otherwise discrete components, while providing the illusion of a glue-less system design.

The architecture of an adaptive manifold can be patterned after the concepts of FPGAs and FPIDs. A typical configuration is abstractly depicted in Figure 13.. The adaptive manifold can be thought of as a complex interconnection network, consisting of a number of electrically isolated wires connected to internal and external terminals. For the purposes of this paper, such isolated wires are referred to as a *net*, and a set of nets is referred to as a *network*. A net itself consists of at least two *terminals*, which represent physically accessible connection points to a net. Internal terminals are usually connected to at least one *switch* contained within a *switchbox*. A switch refers to an individual programmable micro-relay, and a switchbox refers to a collection of such micro-relays, usually all contained within a common package. In Figure 13, the switches are represented as circles, and switch-boxes are rectangles enclosing a number of switches. External, terminals can be thought of as "user input/output" terminals. These are the points of the manifold to which electrical components or connections would be attached. One of the important correspondences between wire-based programmable structures and field programmable devices is the applicability of graph-theoretic treatments, which allow the same heuristics to be applied. Figure 14 illustrates the obvious correlation of nets and switches to nodes and edges, respectively. The nets attached to external terminals are referred to as *terminal* nodes, and other nets are referred as *non-terminal* nodes. The central problem of deciding what switches are to be closed to form the desired connectivity between an arbitrary set of terminal assignments is equivalent to a graph Steiner forest problem, where even the individual graph Steiner tree solution is known to be a non-deterministic polynomial time complete (NPC) problem [64].

As in any reconfigurable system, a concept for controlling the configuration of any switch is required. Once again, it is straightforward to establish such concepts based on the techniques of FPGAs/FPIDs, which often employ a serial configuration approach based on the Joint Test Action Group (JTAG) standard [65]. JTAG, originally conceived as a test / maintenance bus, is ideally suited for configuring field-programmable devices, especially since it is cascade-able, permitting multiple devices to be programmed on the same scan chain. Therefore, a small number of wires in an adaptive manifold must be dedicated to the provision of a configuration bus, which in no other way interacts with the adaptive manifold.

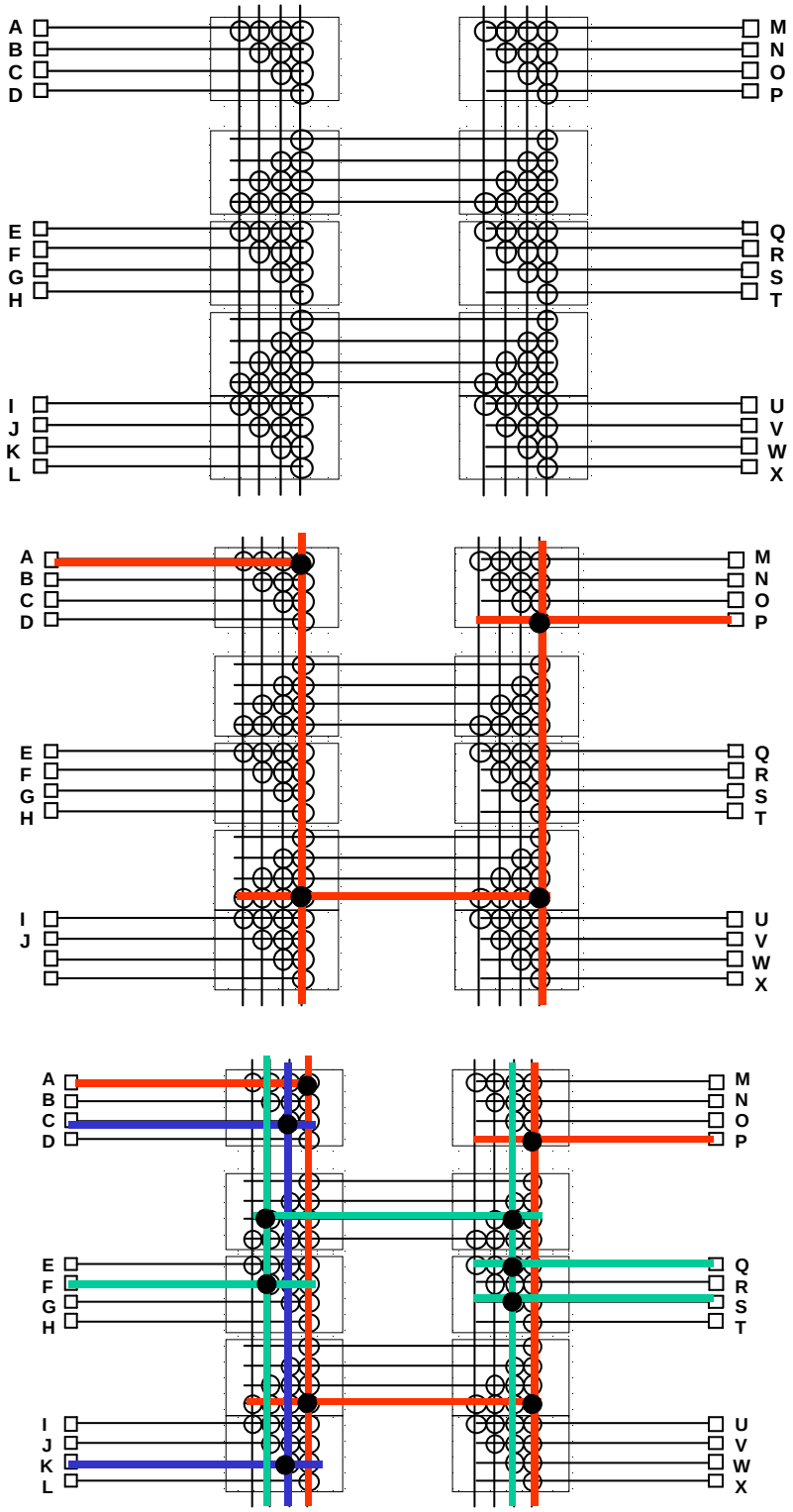


Figure 13. This symbolic representation (top) of the adaptive manifold contains 24 external terminals (A-X), ten small switchboxes, 100 switches, represented. The connection of two or more terminals is accomplished through switch closures. The first example (middle) demonstrates the formation of a net between "A" and "P". The second example (bottom) shows the formation of a network containing nets "AP", "CK", and "FQS".

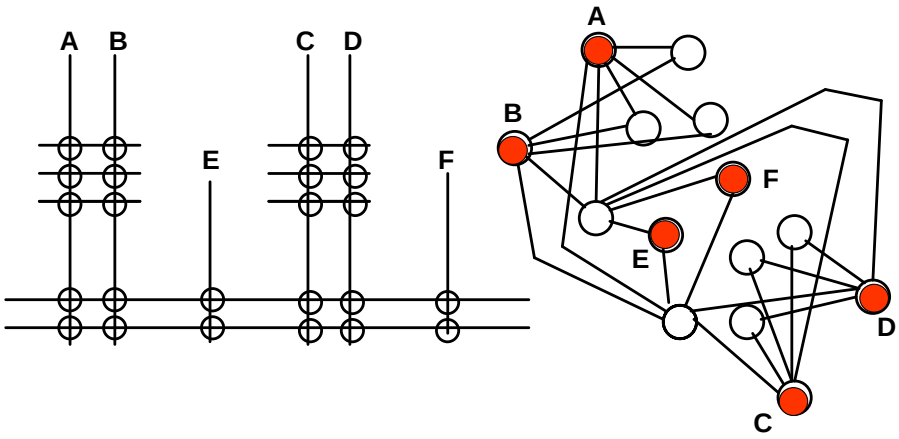


Figure 14. The correlation between nets and switches to graph structures is demonstrated. Nets and switches in the left diagram correspond to nodes and switches in the right graph. Nets A-F correspond to *terminal* nodes, whereas unlabeled nets correspond to *non-terminal* nodes [63].

Relays For Adaptive Manifolds

The desirable properties for relays in adaptive manifolds can be broadly summarized as multi-stable, dense, hot-switchable, and electrically suitable (low contact resistance, high isolation, impedance-matched, high current-handling capability, etc.). The target properties for micro-relays to be used in this application are summarized in Table 1.

For the adaptive manifold, it is necessary to employ switches whose state remains intact until reprogrammed. The vast majority of MEMS switch designs operate on a momentary contact principle, which may result in an unacceptable loss of wiring connectivity upon power removal. Latching MEMS micro-relays based on either thermal [66] or electro-magnetic actuation have been described, and most configurations are bistable, implementing the equivalent functionality of a single-pole, single-throw (SPST) switch. In principle, more elaborate switch configurations are possible (i.e., *m*P*n*T), including more complicated multi-phased arrangements, reminiscent of a washing machine cycle control switch implemented in MEMS structures.

Table 1. Design goals for micro-relays for use in a programmable wiring system.

Design Goals	Logic Switch Manifold	Configuration Switch Manifold	Power Bus Switch Manifold
Constriction resistance	< 10Ω	< 50 mΩ	< 10 mΩ
Configuration	SPST, *m*P*n*T	SPST, *m*P*n*T	SPST
Switch density (#/mm^2)	> 100	> 10	> 1
Energy to switch (0↔1)	< 0.1 mJ	< 1 mJ	< 1 mJ
Hot Switching Capabilities	> 10 TTL/CMOS level loads	10V @ 100 mA	10V @ 1A
Current Handling Capability	> 10 TTL/CMOS level loads	1 A	10 A
Lifetime (cycles)	> 10^7	> 10^6	> 10^4
Time to switch (0↔1)	< 100 μs	< 1 ms	< 10 ms

Density is a fundamental requirement, and it makes mandatory the consideration of micro-electromechanical systems (MEMS) techniques, which can achieve in principle achieve densities beyond 10^4 devices/cm^2. Unfortunately, the requirement for density is in direct opposition to the requirements for low contact resistance and high current-handling capability, due simply to considerations of cross-section area. For this reason, it may be necessary to consider in adaptive manifolds a family

of different size relays and algorithms to ensure that high-current paths are not directed through relays with low current-handling capability.

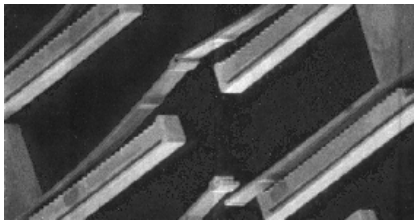


Figure 15. These thermally-actuated bistable MEMS relay based on LIGA fabrication demonstrate a high-performance contact structure for configuration applications. The contacts of two complete relays are shown, the lower programmed in an "OFF" state and the upper in an "ON" state.

In most application regimes except digital and in some cases microwave, minimizing switch resistance is essential. The components of this resistance include the switch constriction resistance, given by [67] $R_{const} \approx \rho/(2a)$ (where ρ is resistivity and a is the effective contact area) and series resistance of interconnections associated with the switch. The effective contact area, which is always less than the apparent area, is based on the accumulation of asperity contact points, which grow with the application of force. Hence, the constriction resistance is therefore a function of applied force F and is given by [68] $R_{const} \approx \rho * F^{-b}$ where $b = 1/3$ in elastic deformation phase and 1/2 during the plastic deformation. In MEMS applications with low constriction resistance, switch operation will likely be in the plastic regime. For typical metals of interest (e.g., Au), a collisional force of at least 100 μN will be needed to generate a constriction resistance of ~ 20 mΩ [70]. Generation of a force in the hundreds of micro-newtons is within the realm of certain MEMS actuators [71].

Comparatively, solid state devices offer tough competition. The source-to-drain resistance of minimum geometry MOSFETS is typically > 50Ω, but it can be substantially reduced by increasing channel width, leading to larger area devices with series resistances below 10 mΩ [69]. MEMS relays are typically much larger, and must have contact resistance properties that surpass those associated with MOSFET structures of a comparable size.

Affecting the reliability of MEMS devices is their tendency to cold weld or bond due to van der Waal attraction. The onset of arcing during switch closure and opening is a significant problem. These problems can result in rapid degradation of the contact surface and a reduction of current handling capacity or destruction of the relay switching capability altogether. The technical issues when the switch operates with applied voltage divide naturally into four groups:

Open switch. In an open macro switch, the minimum in the Paschen curve establishes the minimum standoff voltage. The situation is different for micro switches in that standoff voltage plateaus or decreases as separation decreases [72]. Due to the high electric field gradients developed by nominal voltages, the reduction in effective work-function and tunneling processes dominate and can lead to voltage breakdown below the Paschen minimum. A contaminated surface (e.g., a surface coated with carbon from the decomposition of an organic vapor) can have a breakdown voltage as low as ~40V with a 1 μm gap. The dominant process here is closely related to the processes present in a vacuum discharge [73].

Open → closed. As contacts move together, the breakdown voltage (assuming the open circuit voltage is greater than minimum voltage discussed above) will be exceeded at some point as the gap closes and a discharge will be initiated. If the circuit can deliver sufficient current, this discharge can develop into an arc, which will last until the contacts close completely. This situation is typically not a difficulty for macro switches, but for a MEMS switch it can present a significant problem.

Closed switch. A current flowing in an external circuit will generate a voltage drop across the switch contact resistance given by Ohm's law. Since the current is flowing in only those regions where the surface asperities meet, the power density is quite high and depending on the thermal conductivity of the contact material, can melt part of the contact (for Au, a ~ 0.35V drop can result in melting that part of the surface in contact [74]), potentially resulting in welding the contact surface. The type of load on a hot-swapped [75] relay affects the transient inrush current [76] and hence its lifetime. Resistive loads are the best case situation, as they create inrush currents from up to ten times the steady state value, but capacitive loads can see up to 100X the steady state value. Some loads have dynamic load, such as lamps, in which the initial load is near zero ohms and increases during application of current [77].

Closed→ open. As the switch opens, even if the current flow was small enough to prevent the melting of the contacts, the temporal decrease in contact surface results in increased constriction resistance, increasing the power in the smaller area. It is possible (and normal in macro contacts switching nominal currents) to form a bridge of molten metal between contacts. As the contacts continue to separate, the molten bridge vaporizes and a metal arc discharge is generated [78]. This type of discharge is very destructive to the MEMS switch due to the very small thermal mass available to dissipate the heat.

Practical Work on Relays and Packaging for Manifold Applications

The Air Force Research Laboratory sponsored a series of research efforts, with the eventual goal of creating a practical configuration-grade micro-relay. These efforts date back to the mid-1990's, with a series of graduate thesis efforts at the Air Force Institute of Technology [66,79]. A number of concepts were explored, including the example in

Figure 15, but these early designs were never taken to a higher state of completion.

Beginning in 1998, several contract efforts to create micro-relays were initiated to companies, including Nanosonic, MicroLab (Magfusion), and Los Gatos Research. Of these, Magfusion produced some of the most promising early designs for a complete bistable relay (SPST functionality). Their concept, based on an electromagnetic actuation principle, consistently produced contact resistances below 80 mΩ based on evaluations of dozens of prototypes built in the Figure 17 configuration. Other characteristics of the Microlab switch are summarized in Table 2.

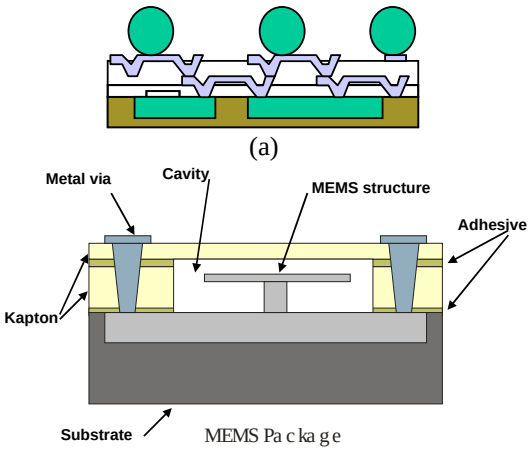


Figure 16. Packaging concepts for a MEMS micro-relay are shown. The first configuration (a) illustrates a compact ball-grid array (BGA) multi-chip module containing several devices. This arrangement can be used to package MEMS die together with conventional CMOS microelectronics, such a driver-controller circuit for a number of MEMS relays. The second concept (b) illustrates the feasibility test structure constructed to examine package and process compatibility issues. This structure can accommodate the significant surface height of the MEMS die (up to 100μ) without compromising interconnect density.

Table 2. Summary of Microlab MEMS relay parameters.

Parameter	Value	Significance
Contact gap (open)	50μ	Sets max switching voltage
Metal thickness	2μ	Series resistance
Contact Velocity	0.2m/s	Affects hot-switching lifetime
Contact Resistance	70mΩ	Device performance
Open/close time	50/200 μsec	Cycle rate

Since the contact resistance measurements were made on bare devices, attention has been turned to creating a packaging concept capable of preserving low contact resistance, low cost production, and compact form factor. An investigation has been performed to examine the feasibility of merge the Microlab switches with the General Electric (GE) High Density Interconnect (HDI) process [80]. Since HDI is based on patterned overlay interconnect, it is possible to implement an efficient direct connection between electrical terminals on the MEMS die surface and a ball grid array (BGA) interconnect, which can lead to a package with very low series resistance [81]. In principle, HDI can be used to form a chip-scale process in which the package structures are formed directly on a wafer surface. Alternately, the HDI process can be used to co-package support components for one or more MEMS die containing a number of micro-relays,an embodiment that is illustrated in Figure 16a. For purposes of test, a simpler configuration was chosen to investigate feasibility, shown in Figure 16b. This test structure was based on the use of a single, thick layer of Kapton dielectric (100μ), followed by a standard thickness layer (25μ). Pockets in the thicker dielectric were formed as a clear zone over areas of the die containing moving structures. Thickness of the dielectric of this initial layer is thus chosen based on vertical clearance requirements. Both layers of Kapton are laminated onto the die surface in sheet form, with thin layers of adhesive. A metallization system is then formed after laser-defined vias are "drilled" to die contact pads, requiring an unprecedented 125μ depth range.

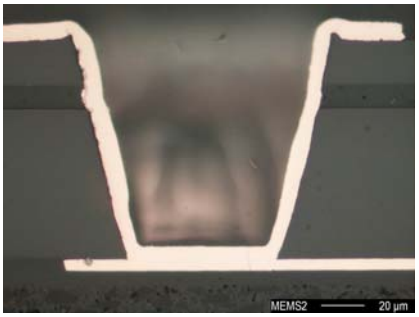


Figure 18. Example of 125μ deep via formed in HDI process (compared to 37μ for standard vias) to contact micro-relay die bond pad to metal level one package metallization in test structure.

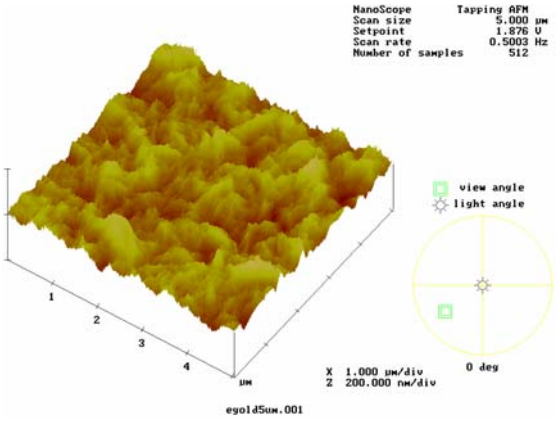


Figure 19. This atomic force microscope (AFM) image of a gold contact surface is based on same metal system used in a MEMS micro-relay, measured in representative use condition.

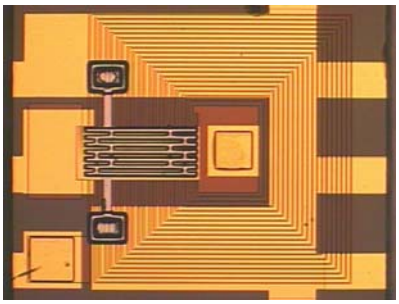


Figure 17. Photo-micrograph of MEMS relay based on electro-magnetic actuation (Microlab, Phoenix AZ).

Results based on a trial fabrication run of several dozen die produced disappointing results from a device yield standpoint, but generated a

number of important insights and demonstrations of extensions of the HDI technique. The deep 125μ HDI vias were successfully demonstrated (

Figure 18), but many test samples were rendered inoperative due to a variety of defect mechanisms. Of these, the most important was a "tent-pole" defect, due to the protrusion of metal as much as 40μ above the MEMS surface in the non-cavity region. Without the clearance of a cavity, the entire dielectric in the area surrounding the defect was lifted, affecting the integrity of the package interconnect through mechanisms including open vias and the unintentional flow of adhesive into the MEMS cavity area. Other problems related to die preparation and handling were encountered, such as die edge roughness and contamination. Each of these defect mechanisms have been ruled as controllable, mostly through more careful die processing and screening. As such, plans are underway for second pass experiments that could ultimately lead to a chip-scale prototype BGA package. Singly packaged Microlab relays are planned in 2002, with eventual plans to form more complex, switchbox arrays, complete with control die, suitable for adaptive manifold applications.

Even as the AFRL program strives to achieve practical MEMS switches for the purposes of programmable wiring, other challenges remain relating to their reliability in general. As an adjunct research activity, AFRL has entered into collaboration with Sandia National Laboratory in an effort to more completely understand the nature of Au-Au contacts. Of primary concern is the evolution of surface morphology over repeated contact cycles in representative use environments. It is hoped that techniques based on extensions of atomic force microscopy (

Figure 19) can be used to provide a more complete understanding of contact erosion, how the erosion might be mitigated, and how the contact resistance of micro-relays might be further improved.

Design Of Simple Smart-Wiring Structures

Though it is believed that a programmable wiring harness will enable unprecedented flexibility in complex systems design and development, no practical work has been done to establish test configurations or to study how the arrangements of wires and switches might be optimized. While the principles of the adaptive manifold discussed in this paper are relatively straightforward, it is felt that constructing simple prototypes would greatly benefit advancement and application of the underlying principles. Just as early FPGA architectures were rudimentary and difficult to apply, it is expected that initial attempts to form adaptive harnesses may experience similar problems. FPGA architectures became powerful and useful, not simply because of the Moore's law density improvements in microelectronics, but also because FPGA architectures became more tuned to the way humans perform electronic design. Even the most advanced FPGAs would be shown as limited to the task of implementing random functions, but they are in fact capable of expressing the majority of human-conceived digital designs (within reason). Similarly, to get the best performance and "expressive range", the adaptive manifold will necessarily undergo an empirical evolution. This evolution will address a range of questions, from the number of long vs. short wires, the quantity of high current relays, the best population strategies for switches in switch-boxes. The way that adaptive manifolds will be applied is another open question. It is expected that the configuration of adaptive manifolds could be compiled, as software programs and FPGA/FPID bitstreams are, in an essentially offline manner. It is also conceivable that on end-use conditions, whether due to a systems failure or an intentional expansion of a system through the addition of more payload multiple configurations could be stored for dynamic, on-line invocation. It is even possible to consider on-line compilation of wiring patterns based functions. These possibilities raise many more intriguing issues, ranging from structural discovery (the dynamic re-mapping of a wiring harness into a graph problem) to the "rules of engagement" for self-compiling, on-line structural modification.

As a first step in understanding a number of these issues, the construction of a simple prototype adaptive manifold has been undertaken. Since the MEMS relays are themselves in development, the first generation adaptive manifold is based on the use of larger "macro-relays". While these larger relays precludes for the most part a practical

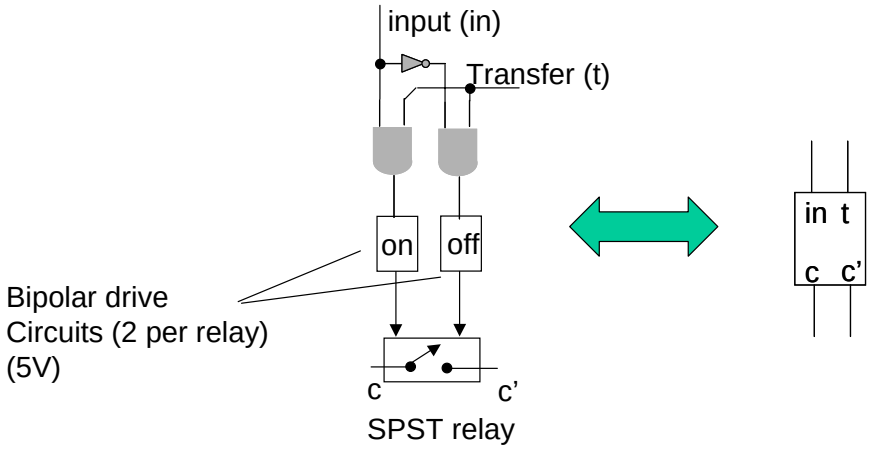


Figure 21. Relay drive circuit.

application of the manifold in an aerospace platform, they will permit a more detailed study of the dynamics in a representative system.

The simple test manifold will be built on a single, large printed wiring board. This inaugural attempt will implement the "tree of meshes" topology, shown in Figure 20. The design can be interpreted as containing a number of switchboxes and a very limited input/output capability (only sixteen terminal nodes). The "tree of meshes" strategy is one of many conceivable routing approaches [82]. Except for fully-populated, non-blocking crossbar topologies, which require $(n^2-n)/2$ switches, most routing strategies are a compromise between expressive capacity and resource utilization. It is very obvious that many possible networks cannot be formed in the Figure 20 architecture, but it is also not certain due to a simple lack of accumulated data which topology would be better.

The design of the manifold configuration network is undertaken independently of the switch / net topology. The relay driver circuit is represented schematically in Figure 21. Since the relay has two stable states, the drive circuit

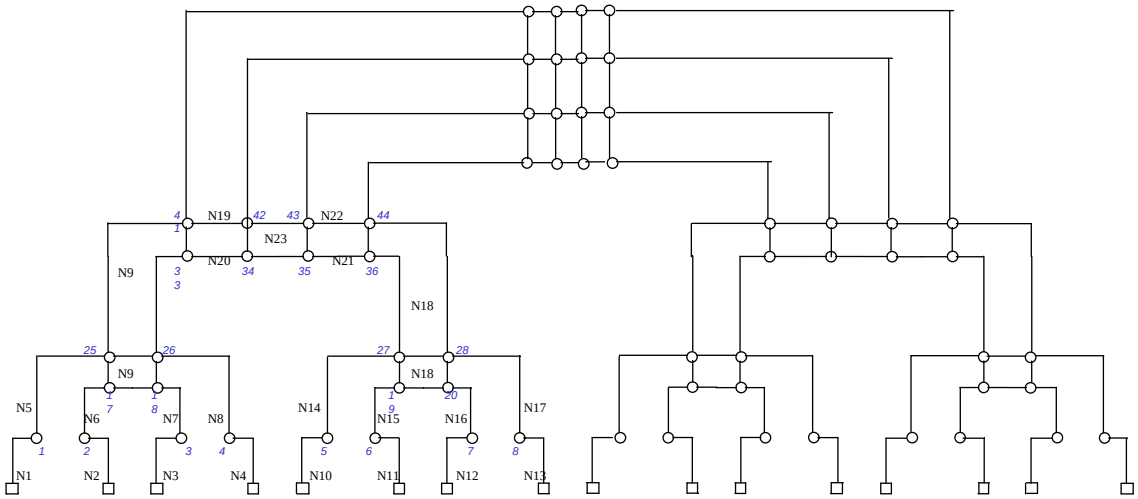


Figure 20. Partially labeled diagram of simple test adaptive manifold, containing sixteen terminal nodes, 36 nets, and 64 relay switches.

must meet three obvious requirements: (1) translation from low-current drive logic at CMOS voltage levels; (2) separate energization for "ON" and "OFF" states; (3) removal of drive current after energization. The Figure 21 circuit achieves this with a programming input and a transfer input, the latter of which gates the energization of the relay coils. In the test board design, most of the required circuit is area accounted for by the discrete drive circuitry and relays themselves.

The overall configuration circuitry is depicted in Figure 22. The synchronous serial configuration approach used is analogous to those used in traditional FPGAs. A JTAG boundary scan concept could be used, but a simpler version is shown for illustrative purposes. In this configuration interface, three inputs (*clock*, *data*, and *reset*) and one output (*done*) are required. To program the relays, reset is asserted prior to the leading (rising) edge of the clock, which latches the value present at the data input. Each subsequent rising edge advances data into the configuration system. A done signal provides feedback on the completion of the bitstream transfer, i.e., it indicates when data for the last relay has been received. The primary difference between the configuration mechanism of a relay-based system and one based on solid state logic is the need to account for the physics and power consumption of the relay energization process. In this design, three shift registers are employed, two in the "forward" direction and one in reverse. The primary forward shift register loads a copy of the desired bit pattern representing the relay program. A second shift register serves as a sentinel, which is a relatively inefficient method of tracking the progress of leading edge of the bitstream. Again, this concept is for illustrative purposes, and in a real implementation, the sentinel shift register could be replaced by an ordinary binary counter modulo *n*, where *n* is the number of relays. In a system involving only one set of relays, the sentinel pulse, which "fires" after the last bit in the bitstream has entered the shift register, triggers a third shift register. This third shift register is clocked by a slow (100 Hz) oscillator. This mechanism insures that: (1) only one relay is energized per cycles, and (2) that the rate of energization is slow enough for the relay coils to physically move the contact from one state to another (for the Microlab relay, the response time is ~ 200μs).

The scheme shown in Figure 22 is flexible enough to permit extension by adding a second copy of a similar structure, connecting the clock and data outputs of the first copy to the clock and data inputs of the second, connecting the "done" signal of the second copy to the "done in" of the first, and tying the reset signals together. In a multiple chain of such circuits, the clock output of the final copy must be connected its "done in" signal. The extensibility of the manifold is important to permit scalability in complex applications

Software. A preliminary software interface for compiling and programming the test manifold has been developed. The TCL-based user interface is shown in Figure 23. The software program accepts user inputs on nodes to be connected and calculates switch closings needed to form connective paths using Djikstra's algorithm [83]. The

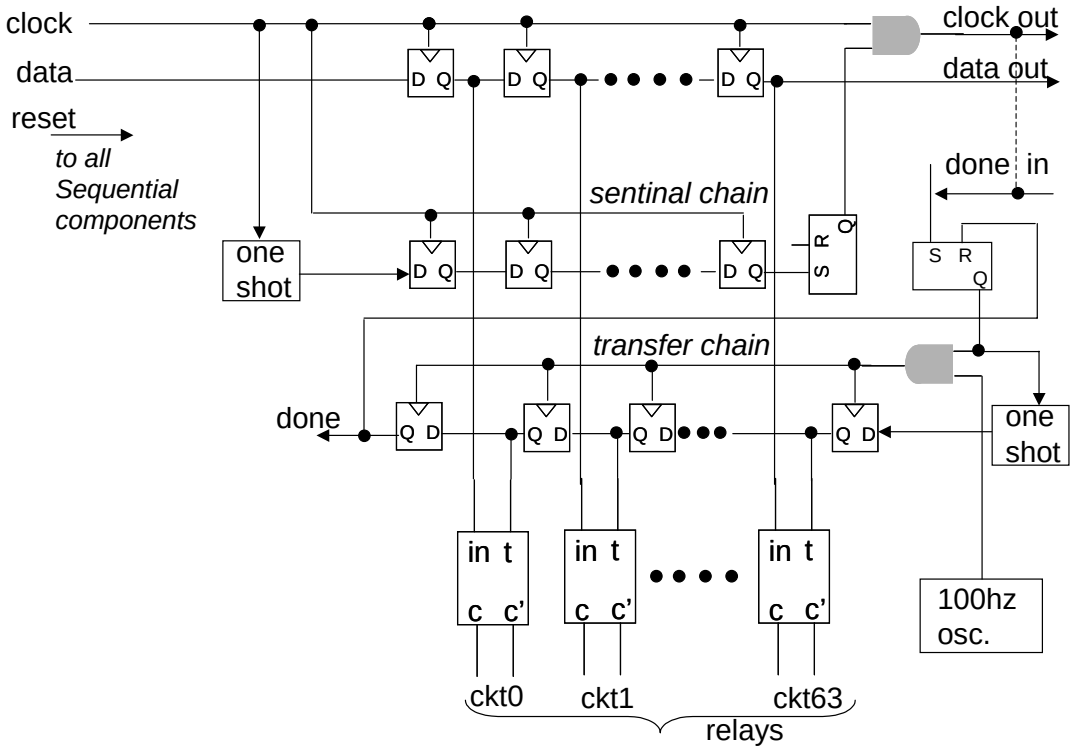


Figure 22. Relay configuration circuit, capable of scaling to multiple sets of relay devices.

simple algorithm is expected to be adequate in small manifolds, but the use of more sophisticated heuristic approaches will undoubtedly be necessary as the application scale grows to hundreds of relays.

Status and Plans. A simple, four-relay test protoboard circuit was used to test the driver circuits and control concepts prior to schematic finalization. At the time of this writing, the schematic capture of a relay test board had been completed, and several copies of a test board are expected to be completed by the end of 2001. The controller state machinery shown in Figure 22 will be implemented in a small FPGA (Altera 10K50) resident on the manifold
board, and a preliminary logic design has been prepared. Test software routines in TCL have been tested on the graph corresponding to the Figure 20 structure, and driver routines have been developed by Utah State University to permit transfer of a bitstream defined by the TCL routines to the adaptive manifold through a parallel port connection on a Windows 2000-equipped PC. When prototypes are populated, testing will be performed first with single copies of the manifold board. Then, the single board will be gradually expanded to a second manifold, chain-connected to the first. Other boards may be connected to extend the configuration further, resources permitting. It will be necessary to alter the graph structures used in the software routines, which were prepared for this eventuality.

After an adequate level of testing, it will be desirable to migrate the large amount of discrete components and FPGA logic to a compact dedicated ASIC, adapted for the MEMS micro-relays described previously. Charge pumps may be included in the monolithic implementation to permit the operation of the relay manifold at reduced supply voltages.

As an exploratory effort, the test manifold is intended to establish a motivation for a range of further investigations of application methodologies, to include the development of benchmarks, the development of embedded compilation approaches, and automated strategies for graph structural discovery and configuration management as the manifold is changed. Developing adequate interconnection densities and arrangements require further study, as had been done in FPGAs as early as a decade ago [84]. Benchmarking, based on some statistical analysis of wiring patterns in avionics, are expected to be useful in establishing switch population strategies and organization of the manifold into zones requiring different types of micro-relays, as suggested in Table 1. As the manifold permits some new concepts in fault tolerance, concepts for detection and circumlocution of defective portions of the manifold or components attached to the manifold remain lucrative areas for future exploration.

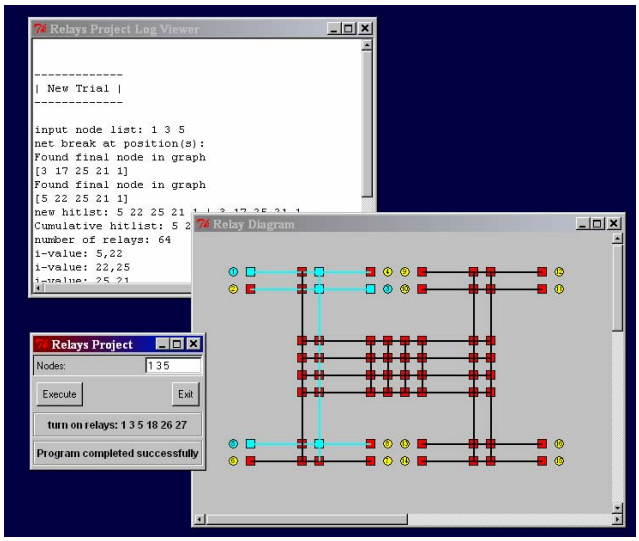


Figure 23. User input interface to software for **compiling and transferring a bitstream to simple adaptive manifold**.

Conclusions

This paper described a concept of fully reconfigurable systems, of which reconfigurable digital approaches comprise a subset of possible techniques. The notion of an adaptive manifold was introduced as binding element of fixed and reconfigurable approaches. In the simplest sense, a reconfigurable or adaptive manifold is a system of interconnections embedded with non-volatile, programmable wires. The states of switches can be related to graph theory problems, and the techniques of FPGAs/FPIDs can be applied to an expanded domain. The limit argument is that aerospace platforms can as a whole be viewed as a complex but fully programmable device. This view of a system permits a number of interesting strategies to be exploited, ranging from rapidly prototyped platforms to more robust forms of fault tolerance. With the adaptive manifold approach, dynamic and on-orbit reconfiguration strategies have a greater reach.

It was then suggested that the fundamental enabler of adaptive manifolds, if not reconfigurable systems, is the micro-relay itself. The requirements of these switches are discussed. As it turns out, most MEMS relays are not suitable for adaptive manifolds, as many are based on momentary contact approaches. A promising bistable design was discussed, along with other efforts associated with making this relay practical for use in aerospace systems.

Finally, a simple test manifold project was described. In a small way, this simple test board, based on macro-relays, becomes a pathfinder in which a range of issues, from extensible manifold configuration to adaptive reconfiguration strategies, can be more fully examined.

The adaptive manifold concept, though straightforward in concept, is a radical concept as a systems engineering technique. Until further investigation is done, it is difficult to quantify how useful the manifold concept will be in actual systems. Undoubtedly, this investigation will need to involve not only the development of test prototypes described in this paper, but more meaningful studies of the wiring harness configurations in aerospace platforms. As in the case of FPGAs, for which benchmarks (MCNC and PREP) were eventually developed, benchmarks will undoubtedly be required for the meaningful advancement of manifolds. Even if the manifold concept is shown valuable from such exploratory research, the final barrier will be the willingness of systems designers to exploit the technology. The success of any technology concept, including adaptive manifolds, will ultimately be decided by the emergence of a user base willing to commit the concept into real-world designs.

4.0 A Single-Asperity Study of Au/Au Electrical Contacts

J. W. Tringe[1], T. A. Uhlman[1], A. C. Oliver[2], and J. E. Houston[2]

[1]U.S. Air Force Research Laboratory, Space Vehicles Directorate
Kirtland Air Force Base, New Mexico
[2]Sandia National Laboratories
Albuquerque, New Mexico

Introduction

The study of the physics of electrical contacts has been an important and productive field since the beginning of the 20th century, and the significance of the microscopic properties of surfaces for defining macroscopic electrical contact properties has long been recognized [85]. With the advent of microelectromechanical systems (MEMS), however, the subject has taken on new life due to the unprecedented low forces and small contact areas available in these devices – both many orders of magnitude below those used in conventional relays. In MEM switches, for example, typical forces are tens to hundreds of μN applied over areas of hundreds to a few thousand square microns. Recently, two latching MEM switches have been reported with very promising characteristics, one with a magnetic actuation scheme [86] and one with a thermal actuation mechanism [87]; both mechanisms use gold-gold contacts because gold is mechanically soft and does not form an insulating oxide. The reliability and performance of these and similar devices depend critically on the nature of the physical contact made by the microscopically rough surfaces in the presence of an applied voltage. Future devices will be even smaller, use lower power and, ultimately, will define and utilize the limiting physical dimensions and electric current levels attainable in a mechanically-formed electric contact.

According to classic constriction-resistance theory, the contact resistance, R_c, through a circular area on a planar, semi-infinite body with constant resistivity ρ can be expressed as:

$$R_c = \frac{\rho}{2a} \tag{1}$$

where a is the radius of circular microscopic contact area. For elastic deformation of the planar, semi-infinite body, $a \propto P^{1/3}$, and for plastic deformation, $a \propto P^{1/2}$, where P is the applied load. Contact resistance for elastic and plastic deformation conditions are given in Eq. 2 and 3, respectively:

$$R_{c,e} = c_e\left(\frac{\rho}{2P^{1/3}}\right) \tag{2}$$

$$R_{c,p} = c_p\left(\frac{\rho}{2P^{1/2}}\right), \tag{3}$$

where c_e and c_p are load-independent constants. Real contact surfaces are not ideally smooth, but are composed of many such constriction-limited contact points, or asperities, and the net result of Eqs. 2 and 3 is that the total contact resistance between two macroscopic rough surfaces is proportional to a root of the applied load – and not proportional to the total macroscopic contact area [85]. Note that this result holds for macroscopic contacts, and

must be used with care in size regimes where the number of contacting asperities is small. Also, although the macroscopic contact area may not directly determine relay performance through contact resistance, the macroscopic area may be important to consider for defining the reliability of the switch since the same microscopic points contacted repeatedly are susceptible to degradation.

Previous large-area contact studies in air have demonstrated how a prepared contact surface can deteriorate with time, and have also shown that a non-metallic contamination film is typically present on metal contact surfaces which produces a resistance in series with the constriction resistance [88]. In fact, this non-metallic "contamination layer," together with other properties of real surfaces that are difficult to quantify, often prevent direct comparison of theoretically predicted behavior with that experimentally observed [89]. For example, a careful study of two microscopically rough gold surfaces in air showed that water and hydrocarbons completely dominate the gold-gold contact properties at low force levels. Current tunneling followed by ohmic behavior was observed as one gold surface approached the other, but even at a contact force of 50 μN the ohmic contact resistance remained several orders of magnitude larger than expected for clean gold surfaces [90]. To obtain repeatable contact results at low forces with resistances appropriate for metals, it is often necessary either to perform measurements in ultra-high vacuum on surfaces cleaned *in situ*, or else to perform violent cleaning procedures in an inert environment. An example of the latter is the so-called "Schaltreinigung" procedure: switching several times with 50 VDC, 50 mA in an N_2 atmosphere immediately prior to making resistance measurements [91].

Experiments using micron-sized probes have been very effective in demonstrating basic properties of metallic electrical contacts. One important study used parabolic gold-coated tungsten probe tips with radii of curvature of a few microns to contact a planar gold surface in nitrogen, through a 2-4 nm thick film of adsorbed hydrocarbons initially present on the gold surface [92]. The authors of this study found that low-resistance, "metallic" contacts (< 200 mΩ) were only observed above an applied force of about 20-60 μN, and further concluded that the affect of the current on contact resistance was dominated by thermal conduction for current levels between 0.1 and 50 mA. In fact, enough thermal energy was generated in these contact events to consistently melt and deform the gold probes at the area of contact, which had a typical diameter of about a micron. At higher forces (e.g., 200 μN), material was observed to permanently transfer in large, several-micron diameter pieces from the tip to the planar contact surface. In these experiments it is likely that the insulating 2-4 nm thick hydrocarbon contamination layer was either displaced mechanically, desorbed or broken down with heat or with electric current, allowing consistent measures of gold-gold contact resistance to be made. Earlier interfacial force microscope (IFM) experiments performed in nitrogen with a tungsten probe tip on electroplated gold at lower forces and currents measured very high (tens of thousands of ohms) contact resistance due to the contamination layer or oxide on the tungsten tip. Under certain force and voltage conditions, this electrical resistance was observed to drop abruptly in coincidence with movement of the probe toward the surface, implying mechanical as well as electrical breakdown of an insulating layer [93].

In an important ultra-high vacuum (UHV) study, a tungsten probe with radius of curvature of about 1 μm was brought into contact with clean and oxidized Ni(111) surfaces while tip-surface forces and contact resistances were simultaneously measured[94]. This study demonstrated that at relatively low applied loads (1-30 μN), contact resistances on the order of 1-2 Ω can be achieved for clean Ni(111) surfaces with the tungsten probe; with two monolayers of oxygen on the Ni(111), the contact resistance increased approximately three orders of magnitude. Further, it was shown that adhesion forces between 10-20 μN were consistently present for applied loads between 1 and 100 μN. With two monolayers of oxygen on the Ni(111) on the surface, the adhesion force was reduced by a factor of two to three.

Repeatable low-contact resistances at relatively low forces in MEM relays demonstrate that the surface contamination layer can be managed under certain conditions. For example, applied loads between 50- 200 μN give contact resistances of less than 100 mΩ in a magnetically-actuated relay with gold-gold contacts, for current levels of 240 μA and above, to greater than 5×10^6 cycles [86]. In a thermally actuated switch with gold-gold contacts, forces of about 20 μN give contact resistances in the range 2-36 Ω under loading of 1 mA for more than 10^8 cycles [87] Ultimately, the reliability of these and similar devices are limited by the effects of the contamination layer, possibly in association with changes in the rough topography of the contact surface.

In this work, we use an IFM scanning force-probe technique to examine single-asperity contacts with an electroplated gold surface in dry nitrogen. The role of the surface contamination layer is examined explicitly, and forces and currents on the order of those applied to single asperities in MEM devices are observed.

Experimental Methods

The experimental apparatus used for this work was the interfacial force microscope (IFM), details of which are given elsewhere [95]. Briefly, an electrochemically-sharpened probe tip is fixed to a self-balancing force sensor which allows both attractive and repulsive forces to be measured dynamically as the tip is brought into contact with a surface. Forces from about 0.01 to 100 μN can be measured. For the experiments reported here, the tip was grounded and a voltage was applied to the electroplated gold substrate with a constant-current source capable of generating currents from 0.01 to more than 1 mA. An environmental control chamber was mounted around the measurement apparatus and flooded with dry nitrogen for the experiments. Temperature and relative humidity were continuously monitored; all experiments were performed at about 23 C with relative humidity less than 3%.

An electrochemically-sharpened parabolic tungsten tip used for these experiments was coated with gold: immediately following hydrofluoric acid dip and de-ionized water rinse, the tip was cleaned by argon plasma (50 W, one minute), and 50 nm of gold was deposited on the tungsten tip by sputtering. In Figure 24, scanning electron microscope (SEM) micrographs show the smooth, gold-coated probe surface in profile at high and low magnifications; the tip has a radius of curvature of about 800 nm. The electroplated gold substrate, a few microns thick, was made using methods essentially identical to those used to make the magnetically actuated MEM relays reported elsewhere [86].

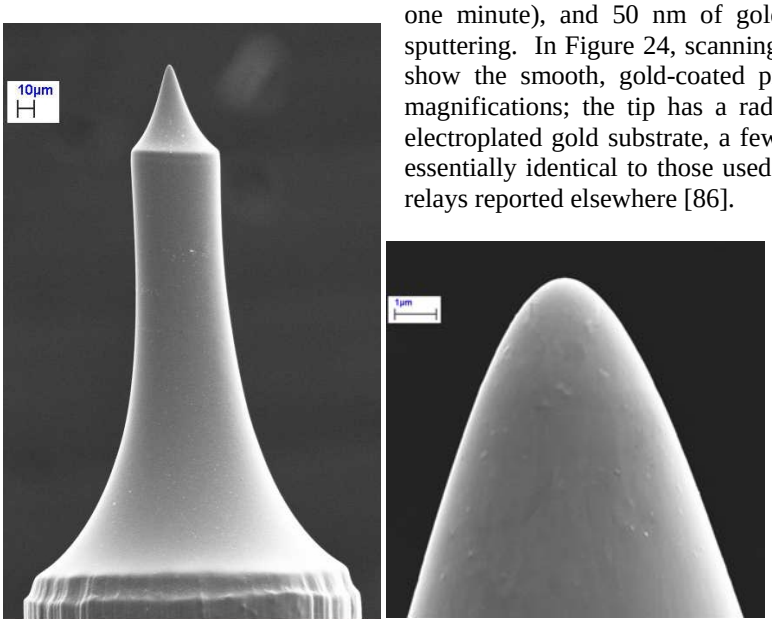


Figure 24. SEM images of probe tip used for experiments reported here; 50 nm of gold was sputtered on a parabolic tungsten tip with 800 nm radius of curvature.

In a typical IFM force profile experiment, the tip was brought toward the substrate at a rate of about 4.2 nm/s. The tip's approach continued after contact until a preset repulsive force was reached, then the tip withdrew at the same rate. During the entire contact event the attractive and repulsive forces on the tip were measured, together with the applied voltage on the tip from the constant current source. Since the current was maintained constant during the contact event, the probe-substrate contact resistance could immediately be found from the measured voltage value divided by the current level. Due to an offset in the constant-current source of 0.5 mV, there were apparent resistances of approximately 50, 5, and 0.5 Ω in series with the tip-substrate resistance for constant-current levels of 0.01, 0.1 and 1 mA, respectively. The maximum applied voltage between the tip (ground) and substrate was 5.6 V at all current levels; the maximum measurable contact resistances were therefore 56 000, 5 600, and 560 Ω for constant-current levels of 0.01, 0.1 and 1 mA, respectively.

Figure 25 shows a representative IFM profile with normal force and contact resistance plotted as a function of relative displacement between the tip and substrate. Repulsive forces are shown as positive and attractive forces are negative. The nominal point of zero separation, or first physical contact, between tip and substrate can be taken as the relative displacement distance, d, at which the tip first begins to deviate from its increasingly attractive behavior. This point appears at about d = 13 nm in Figure 25. The maximum attractive force is negligible on the scale shown, and the minimum contact resistance is about 9 Ω for a constant current of 0.01 mA. Current flow begins at a relative displacement of 9 nm when the tip experiences a repulsive force of about 6 μN; this repulsive force is subsequently referred to here as the "breakdown" force. The hysteresis evident between the force-displacement curves measured as the tip approaches and withdraws indicates that some sort of plastic deformation occurred during the contact event.

A parabolic tip with a radius of curvature of 800 nm intersecting a planar substrate to a depth of 20 nm defines a contact area of about 0.1 μm^2. Assuming this contact geometry, the maximum 40 μN force applied during these

experiments generates a stress of about 0.4 GPa. Because the real electroplated gold substrate is rough, the actual area of physical contact between probe and substrate is less than estimated with the planar substrate assumption, so that the actual stress is higher than 0.4 GPa. During the experiments reported here, the gold-coated tungsten tip did not plastically deform and no material transfer from or to the tip was observed, as demonstrated by SEM images taken before and after contact events.

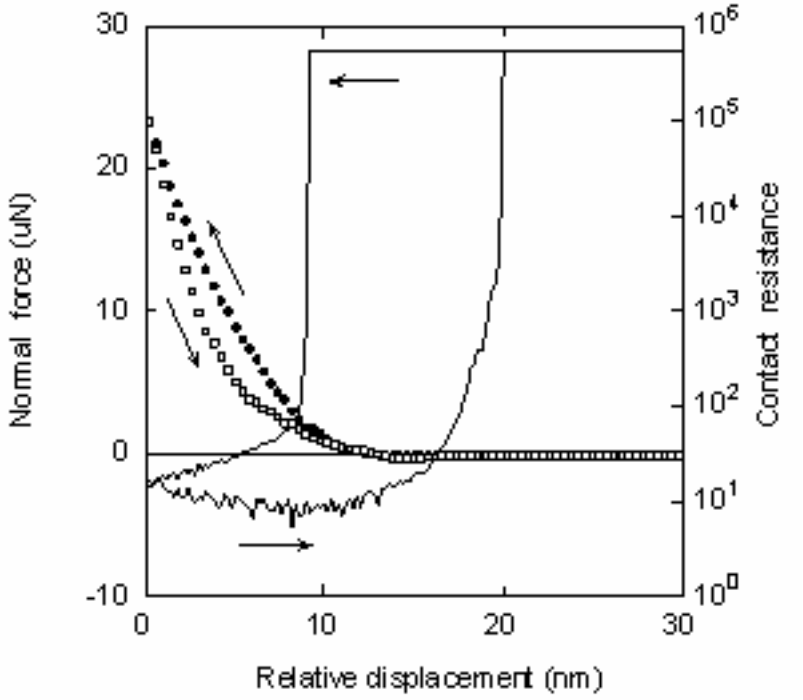
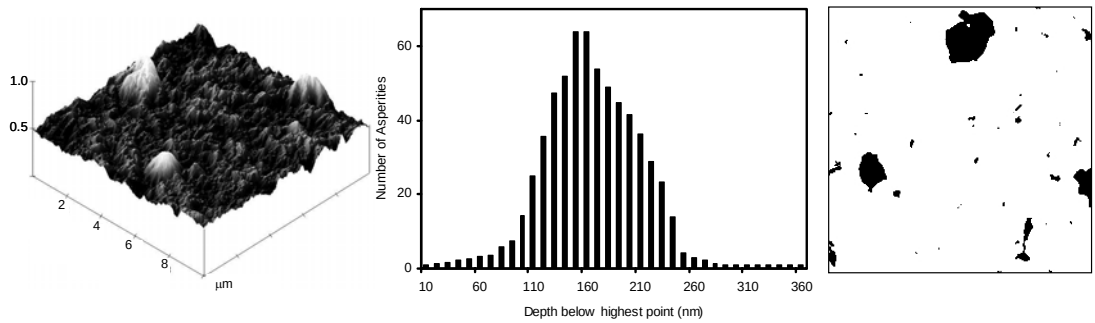


Figure 25. A representative IFM profile where contact resistance and force applied to the tip are given as a function of the relative displacement of the tip. Measurements are made in a nitrogen ambient at room temperature. Positive forces on the left axis are repulsive, and negative forces are attractive. The resistance is saturated at 560 kΩ when the tip is not in electrical contact with the substrate; when electrical contact is made, the current is maintained constant at 0.01 mA. Current flow continues more than 5 nm past the point of initial contact as the tip withdraws from the substrate.

Results

In order to understand the nature of electrical contact between rough surfaces, it is useful to observe the topography of the surfaces in question in order to estimate what types of features and areas might be active in an electrical contact event. A 10-μm × 10-μm atomic force microscope (AFM) image of an electroplated gold surface is shown in Figure 26a. This image is taken from the same gold film used for contact experiments reported here. Averaged over five 10 μm × 10 μm areas, the RMS roughness of the surface is found to be 34.5 ± 5.5 nm, and there are 3.6 ± 1.5 features in this total scan area with a height greater than twice the RMS roughness. These features would act as first conduction paths if two similar electroplated gold surfaces were brought into physical contact with an electrical potential between them. The average combined area of these highest features represents only about 1% of the total 100 μm^2 area. A histogram with the number of asperities as a function of depth relative to the highest point in a 10 μm × 10 μm AFM scan is given in Figure 26b, averaged over five 10 μm × 10 μm scans. The number of distinct asperities increases to a depth of 160 nm, after which the asperities begin to merge. A section of the 10-μm × 10 μm scan shown in Figure 26a is shown in Figure 26c; this section was created at a depth of 100 nm above the lowest point in the scan, and individual asperities are highlighted. It is evident that a few asperities at this depth are spread across the 10-μm × 10-μm area, but that the three highest asperities dominate the topography.



(a) (b) (c)

Figure 26. (a) A 10 µm×10 µm atomic force microscope image of the electroplated gold substrate; the surface features are dominated by three maxima. (b) A histogram with the number of asperities as a function of depth relative the highest point in a 10 µm × 10 µm AFM scan, averaged over five 10 µm × 10 µm scans. The number of distinct asperities increases to a depth of 160 nm, after which the asperities begin to merge. (c) A section of the 10 µm × 10 µm scan shown in Fig 3a; this section was created at a depth of 100 nm above the lowest point in the scan, and individual asperities are highlighted.

In a first set of IFM experiments, the substrate was prepared by rinsing in de-ionized water for about two minutes, drying in dry N_2 for about two minutes, and finally was kept in a flowing dry N_2 environment for 36 hours (relative humidity < 1%). Force-resistance profiles were measured for constant current levels of 0.001, 0.01, and 0.1 mA at eight different locations on the electroplated sample surface with applied loads of up to 31 µN; multiple contact measurements were made at most locations. Table 3 summarizes the minimum average contact resistance at each location and current level, together with the maximum force applied. In general, the minimum contact resistance for these experiments was on the order of tens of thousands of ohms, and in some cases several orders of magnitude higher. Electrical contacts were not consistently made, in spite of high applied forces, and when contacts were made there was a great deal of resistance variability between contact events – even those made in nominally the same location on the substrate. A typical force profile is shown in Figure 27(a) for a constant current contact event with 0.01 mA. The point of first contact appears at d ≈ 15 nm, and the maximum attractive force experienced by the tip during the contact event is negligible on the scale shown. Electrical contact begins at a repulsive breakdown force of approximately 1 µN, but the initial contact resistance is about 10,000 Ω and does not decrease much as the applied force increased to 12 µN. Upon withdrawal, the contact resistance remains at ~20,000 Ω to a relative separation of d ≈ 30 nm extending to ~15 nm from the surface, implying the formation of a plastic conduction path.

Table 3. Summary of pre-ozonation contact events at eight locations.

Spot	# electrical contacts /total attempts	Current (A)	Max F (µN)	*Avg R (kΩ) (with standard deviation)
1a	4/4	$1x10^{-6}$	2.4	276±485
1b	6/6	$1x10^{-6}$	10.9	330±586
	3/3	$1x10^{-5}$	11.9	123±87
	1/5	$1x10^{-4}$	20.3	20 (one meas.)
1c	4/5	$1x10^{-4}$	31	21.0±4.2
1d	1/5	$1x10^{-4}$	24	20 (one meas.)
1e	5/5	$1x10^{-4}$	29	13.5±9.6
1f	2/6	$1x10^{-4}$	31	21.5±2.7
1g	0/4	$1x10^{-4}$	21	N/A
1h	2/6	$1x10^{-4}$	23.5	22.5±3.5

* Average of contact events with successful electrical contact

Mechanically, the electrical contact system consists of the tip, the substrate, and the intervening contamination layer. The nominal 'elastic modulus' of the tip-contamination film-substrate system was obtained by applying the Johnson-Kendall-Roberts (JKR) contact mechanics model to the high-force part of the force-displacement profile where hysteresis was minimized [96]. The JKR model is a continuum theory of contact between a solid sphere and an infinite planar substrate; it is an extension of Hertzian contact theory and takes into account the additional increase in the contact area resulting from the adhesive forces between the two surfaces. The theory allows a direct estimate of the surface free energy of the interface as well as the work of the adhesion between the solids. The JKR fit is shown by a solid line in Figure 27 (b) and gives a composite modulus of 23 GPa. In general, similarly-estimated moduli range from 5 to 37 GPa.

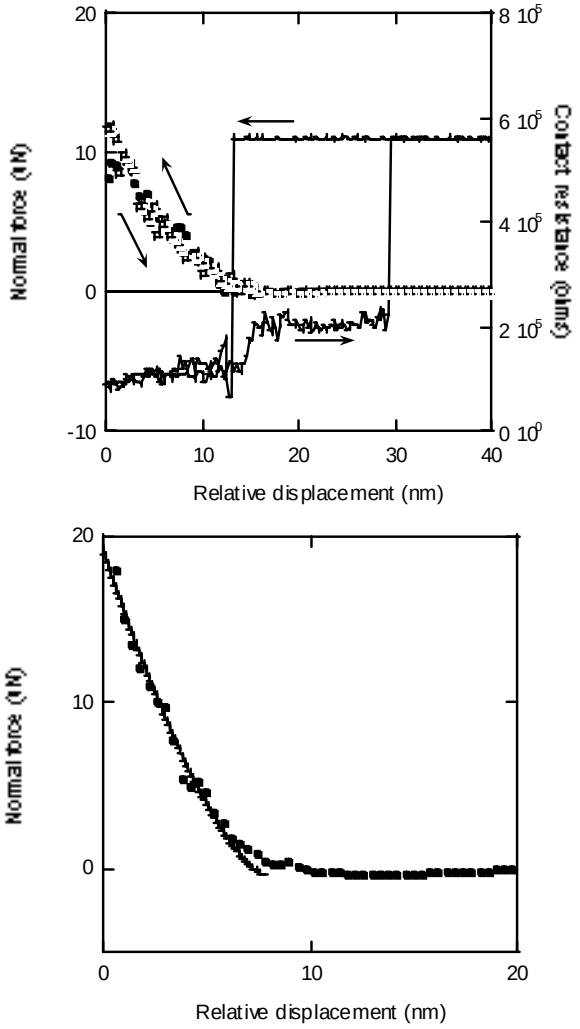


Figure 27. (a) A high-resistance, 0.01 mA contact event prior to tip and substrate ozonation. The minimum contact resistance is about 10,000 Ω, even at an applied force of 12 μN. Current flow continues more than 10 nm past the point of initial contact as the tip withdraws from the substrate.(b) JKR fit (solid line) to the force profile indicates an elastic modulus of 23 GPa.

In a second set of IFM experiments, the tip and substrate were exposed to ozone for 40 minutes [97], then immediately placed in the aforementioned experimental environment. Force-resistance profiles were taken at constant current levels of 0.01, 0.1, and 1 mA at eighteen different locations on the substrate with applied loads of up to 36 μN. Multiple contact measurements were made at each location. Table 4 summarizes the average minimum contact resistance at each current level, together with the maximum force applied. Accounting for the

voltage offset in the constant current source, the minimum contact average resistance is less than 10 Ω for all three current levels. After ozonation, contact measurements were quite consistent, as evidenced by the low standard deviations in resistance in Table 4. Note that contact locations are labeled according to the order in which they were probed, so that "2a" was the first spot, "2b," the second, etc. Contact measurements at 0.01 and 1 mA were made alternately at different locations at the end of the experimental series, and resistance values at both current levels remained consistent with previous measurements. A breakdown force was observed at many, but not all, contact locations following ozonation. At each location in Table 4, the repulsive turnaround force was started at 2 μN, and gradually increased with successive contact measurements until a low contact resistance was observed. A histogram representing a typical measurement progression is shown in Table 5, where the breakdown event was observed at a force of about 28 μN. In some measurement locations, however, the low contact resistance state was observed when the probe tip first experienced very small (<0.1 μN) repulsive, or even attractive forces on the approach.

Table 4. Summary of post-ozonation contact events at eighteen locations.

Spot	# electrical contacts /total attempts	Current (A)	Max F (μN)	*Avg R (Ω) (with standard deviation)
2a, 2b, 2c, 2d, 2e. 2f, 2p	67/99 (68%)	$1x10^{-5}$	36	59.0±19.9
2g, 2h, 2i, 2j, 2k, 2l	48/60 (80%)	$1x10^{-4}$	24	9.0±2.2
2m, 2n, 2o, 2q, 2r	38/44 (86%)	$1x10^{-3}$	32	3.0±.7

* Average of contact events with successful electrical contact

A typical force profile from a post-ozonation contact experiment is shown in Figure 28a for a current of 0.1 mA. The breakdown force needed to achieve the low-resistance state of ~9 Ω was about 5 μN. The plastically deformed conduction path, which is evident during the withdraw stage in Figure 25and Figure 27 (a), does not appear in Figure 28a. However, in general this conduction path did appear in the post-ozonation contact measurements to an extent which was comparable with the pre-ozonation contact events. Similar to the non-ozonated data, only small attractive forces are observed during either approach or withdrawal stages in the profile in Figure 28a. The nominal 'elastic modulus' of the tip-contamination film-substrate system obtained by applying JKR analysis, as shown by the solid line in Figure 28b, is 20 GPa.

Table 5. Maximum force and minimum resistance contact event history for one location following ozonation.

Iteration	Contact Resistance (Ohms)	Normal Force Maximum (uN)
1	∞	2
2	∞	4
3	∞	8
4	∞	16
5	∞	24
6	3	32
7	3	32
8	3	32

X-ray photoelectron spectroscopy (XPS) and time-of-flight secondary ion mass spectrometry (TOF-SIMS) characterizations were performed on gold samples from the same wafer as the tested gold from experiments above,

both before and after exposure to ozone. XPS spectra revealed gold, oxygen, nitrogen, carbon and small amounts (< 1 At%) of sulfur to be present in the near-surface region of four separate electroplated gold samples. Following ozonation, the C_{1s} XPS spectra changed slightly, possibly as a result of changes in C-O chemistry, but the major peak in this spectrum was due to adventitious hydrocarbons. Tougaard analysis on the spectra suggests the carbon-containing layer has a thickness of 4-6 nm. With TOF-SIMS, $Au_x(CN)_y$, Au_xI_y, SO_3^- and SO_4^- were observed before exposure to ozone. Following exposure, however, the $Au_x(CN)_y$ and Au_xI_y concentrations were reduced (Au_xI_y signal was absent), while SO_3^- and SO_4^- were concentrations were increased.

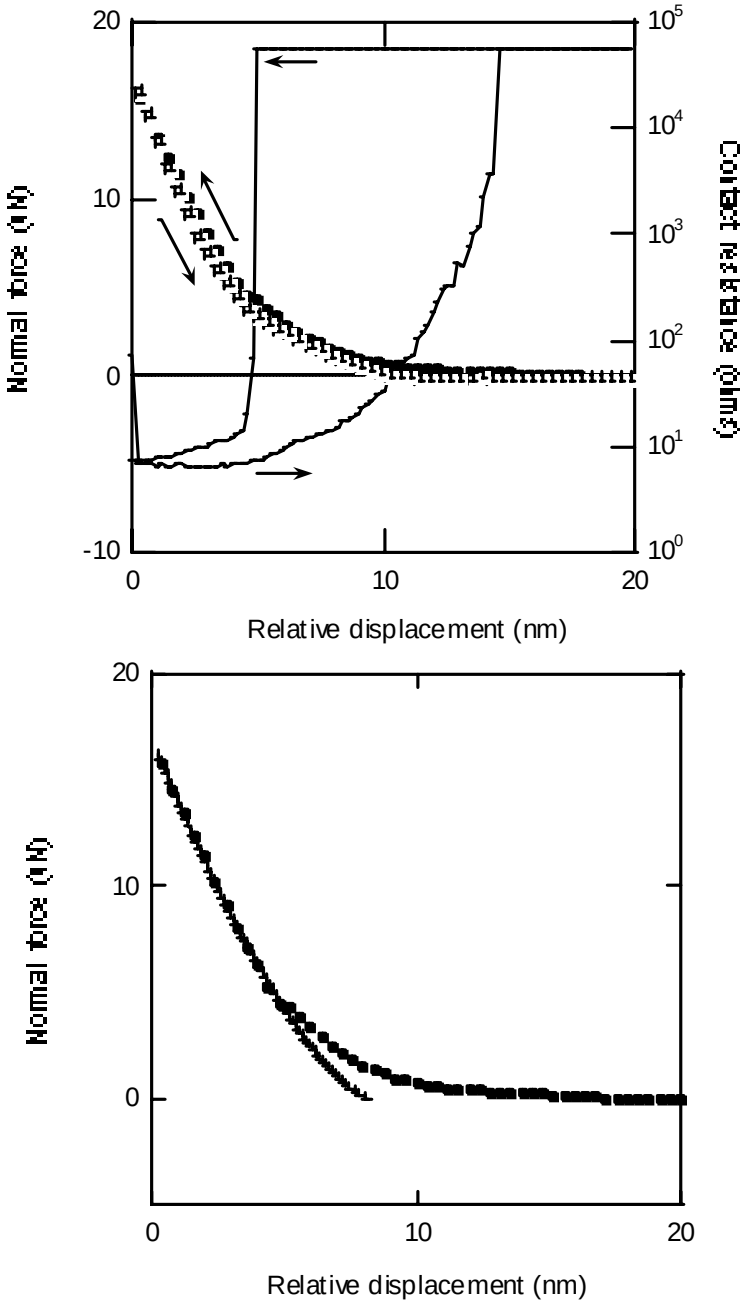


Figure 28. (a) A low-resistance 0.01 mA contact event following tip and substrate ozonation; the minimum contact resistance is about 9 Ω, achieved after a repulsive breakdown force of ~5 μN is applied. (b) JKR analysis of the force profile indicates an elastic modulus of 20 GPa

Discussion

In this work, we use an IFM scanning force-probe technique to examine single-asperity contacts with an electroplated gold surface in dry nitrogen at room temperature. The role of the surface contamination layer is examined explicitly, and forces and currents on the order of those applied to single asperities in MEM devices are observed. A previous study [92] used a micron-sized probe to perform a similar experiment, but with significantly higher forces and currents, comparable to total forces and currents used in MEM relays with much larger contact areas. In an actual relay, however, these forces and currents are distributed over larger areas and over many asperities, so that realistic force and current per asperity is significantly lower. For example, in a study of topographical changes on 1 μm-thick cobalt-hardened electroplated gold contact surfaces caused by loading (without current), it was concluded that 80%-90% of the load generated by smooth ruby spheres was carried by elastically-deformed asperities or asperities plastically deformed less than 30 nm [98]. The ruby spheres had radii which varied between 790 and 3180 μm, and a force of 9.8×10^5 μN was applied to generate stresses up to 6.6 GPa over a 150-μm^2 area at a (planar) penetration depth of 30 nm. Here it is evident that even very large stresses are supported by many asperities in electroplated gold films.

AFM images (e.g., Figure 26a) of the electroplated gold surface used both for this study and for magnetically actuated devices [86] show a few asperities higher than twice the RMS roughness which support force and conduct current in the contact event. Over an area of 500 μm^2, 18±8 such asperities were observed by AFM, implying that the total force and current in a MEM devices with an active contact area 2000 μm^2 may be distributed over about 72 asperities. Assuming 72 active asperities in a 2000 μm^2 contact area in a MEM switch operating with 100 μN force and 240 mA current, each asperity would be subject to a force of ~1.4 μN and a current of about 3.3 mA. This is consistent with the resistance values obtained in this study for post-ozonation conditions, since single-asperity resistances on the order of 3-10 Ω were measured at 0.01, 0.1 and 1 mA. Seventy-two asperities in parallel, each with a resistance of 5 Ω, give a composite surface contact resistance of 69 mΩ, comparable to the < 50 mΩ value reported [86]. Note that for a metallic contact between a probe tip with 1-μm radius of curvature and an atomically flat, clean surface in UHV conditions, resistances of 1-2 Ω were recorded [94]. Much smaller resistance values, on the order of 100 mΩ, obtained with a micron-sized probe reported elsewhere may be due to significantly increased contact areas induced by large forces and currents [92]. Contact resistances in that study were measured for forces between 20 and 500 μN with currents between 0.1 and 50 mA. For those experiments, the actual contact diameter was estimated to be between 0.8 and 3 μm based on damage to the probe visible by SEM following contact events.

Both before and after exposure to ozone, the composite elastic modulus of the tip-contamination layer-substrate system as determined by JKR analysis of the IFM force profiles was significantly lower than the elastic modulus of bulk gold, 78 GPa [99]. For example, in pre-ozonation contact locations, the composite moduli determined from force profiles ranged from 5 to 37 GPa. Post-ozonation moduli were comparable, as shown in Figure 28b. From these low composite modulus values, it is inferred that the modulus of the contamination layer is dominating the mechanical interaction between tip and electroplated gold since the gold-coated tungsten tip is significantly stiffer than the electroplated gold[92]. Although some hysteresis is often observed in the force profiles, the JKR analysis was consistently applied to the higher-force regime of the force-versus-displacement curve where this hysteresis is minimized. Therefore, the moduli obtained from the JKR analysis are reasonable estimates of the actual properties of the mechanical system under observation, and Figure 25 and Figure 28a show the effects of a water-hydrocarbon contamination layer, on the order of 5 nm thick, being deformed by the tip as it approaches. The forces used for this study are too low to appreciably deform significant volumes of the electroplated gold [98], and the tip itself showed no evidence of deformation or damage at the conclusion of the experiments by SEM. Other investigators used an uncontrolled current source with an electrochemically-sharpened gold tip and a single-crystal gold substrate, and found that currents between 0.015 mA and 1.5 mA resulted in quantized conductance and in the creation of gold nanostructures on the substrate at atmospheric pressure and room temperature. The minimum contact resistance measured was less than 100 Ω[100]. Neither quantized conductance nor gold nanostructures were observed in the experiments reported here, possibly because of the dominant role of the contamination layer on these rough electroplated gold substrates.

Electrical contacts in ultra-high vacuum (UHV) demonstrate dramatically contrasting, but complementary behavior to contacts made in air or nitrogen. Figure 29 shows two UHV first-contact events between a clean parabolic tungsten tip and a clean single-crystal gold substrate taken at room temperature under constant-current conditions of

1 and 10 µA with a probe having a radius of curvature of about 400 nm[3]. The 1 µA contact of Figure 29a indicates only small attractive electrostatic forces before contact, after which the force suddenly jumps to an attractive value near -50 µN and the contact resistance suddenly falls to just above 10 Ω. As the tip continues its approach and the force rises toward repulsive values, the contact resistance slowly decreases to a minimum value of about 6 Ω, which is approximately the floor of the resistance measurement. It should be emphasized that during this portion of the approach curve the behavior is totally the result of very strong metal/metal adhesive forces, and that the stable sensor is maintaining the tip/sample separation. In the absence of this stability the two surfaces would crash together to the point of equilibrium, i.e., the zero of force.

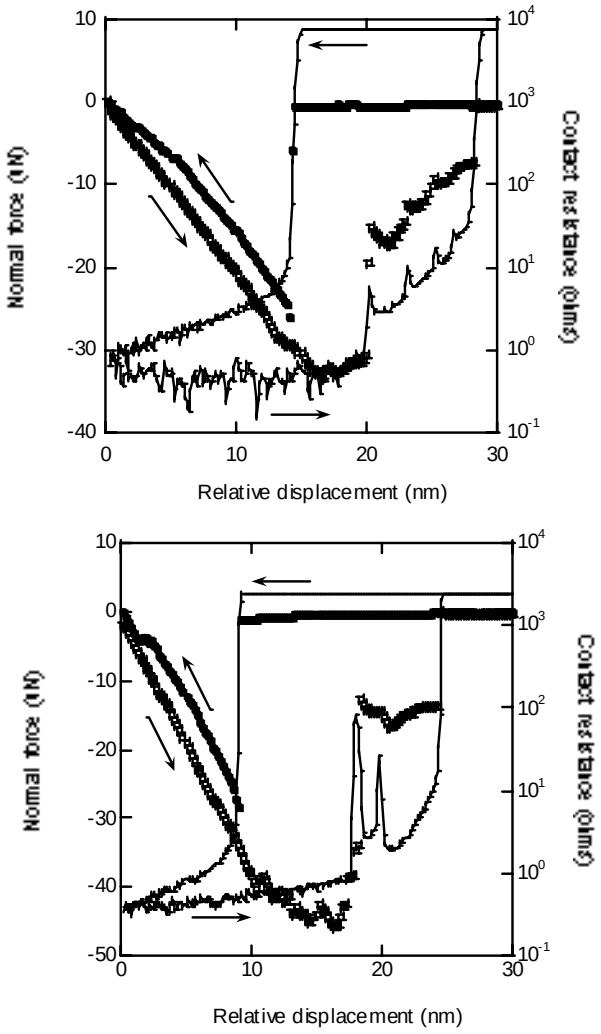


Figure 29 (a) A UHV first-contact event between a clean parabolic tungsten tip and a clean single-crystal gold substrate at room temperature for a constant current of 1 µA.17 The initial contact is characterized by extremely large attractive forces (~50 µN) that are instantly developed as the low-resistance contact is realized; the minimum contact resistance is less than 10 Ω. As the tip withdraws, the contact area remains reasonably constant within the range of the stated resistance floor until about 2 nm beyond the initial contact. At this point, the force flattens as the contacting material plastically flows and the contact resistance begins to slowly rise. (b) A UHV first-contact event between the tungsten tip and the clean single-crystal gold substrate for a constant current of 10 µA..

[3] These difficult UHV IFM measurements were performed by H. Cabibil and J. E. Houston at Sandia National Laboratories.

The reduction in contact resistance in the attractive approach region is the result of the increasing contact area. The fact that the force does not follow a normal Hertz relationship, along with the appearance of a hysteresis loop upon withdrawal, indicates that the interaction is a combination of elastic and plastic compliance. In addition, the width of the loop indicates a plastic "indentation" having a depth of ~2 nm. As the tip withdraws, the contact area remains reasonably constant (within the confines of the instrumental resistance floor) until about 2 nm beyond the initial contact. At this point, the force flattens as the contacting material plastically flows and the contact resistance begins to slowly rise. At a relative displacement near 20 nm, the material suddenly yields, decreasing the attractive force by about 40% and accompanied by a small jump in contact resistance. Both of these relaxations heal slightly (probably due to Au surface diffusion) and are followed by a series of similar relaxation events until final fracture near 28 nm. A very similar behavior is seen for the 10 μA profile of Figure 29b, although the higher current has reduced the number of minor yield events and the plastic flow is extended due to an increase in surface diffusion. Similar behavior is observed by scanning tunneling microscopy (STM) under UHV conditions at liquid helium temperatures, when connective "necks" of gold are found to plastically deform and induce step-wise changes both plastic yield and conductance. However, adhesion forces are much lower (<2 μN) with the STM contact geometry at these low temperatures [104].

These results clearly indicate that even if a UHV MEMS relay could be fabricated, it is unlikely that it could be opened once it had closed. Hundreds of μN are typically available in MEM actuation mechanisms to close and open the contact. However, if under UHV conditions each individual asperity develops ~50 μN of attractive force, and there are many tens of asperities active during a relay contact event, the force necessary to break the contact would be prohibitive. In contrast, the attractive forces in the nitrogen-ambient experiments are consistently small during the extended conduction stage of the contact event, as shown in Figure 25, Figure 27a and Figure 28a. Here the contamination layer, composed mostly of hydrocarbons, is preventing the large gold-gold adhesion forces from forming. The reduction in adhesion force with modification of a clean metallic surfaces is consistent with a trend observed previously in UHV experiments where two monolayers of oxygen on a Ni(111) on the surface reduced the adhesion force on a tungsten probe by a factor of two to three relative to the force generated on the same probe by a clean Ni(111) surface [94]. The contamination layer on the gold surfaces in experiments reported here also acts as a plastically-deforming conduction path. Given the similarity in the magnitudes of the resistances observed in UHV and nitrogen environments, the contamination layer seems to be acting as an excellent conductor, as well as a surface passivator. Although the layer may ultimately limit the lifetime of a gold contact in air or nitrogen, without such a contamination layer gold-gold electrical contacts would likely weld closed immediately and permanently.

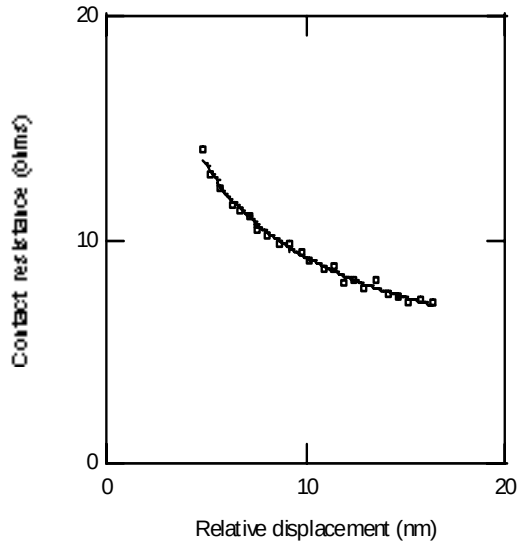


Figure 30.) A power-law fit to the contact resistance plotted as a function of applied force yields an exponent of 0.5, in good agreement with the plastic deformation behavior predicted by the constriction resistance model. However, many contact events exhibited significantly lower exponents than predicted by the model for either elastic or plastic contact behavior.

As discussed above, the contact resistance decreases with increasing contact force as qualitatively predicted by the constriction resistance model. For example, in Figure 29a, there is a ~5 nm region during the approach stage over which the tip experiences a measurable contact resistance, which slowly decreases with increasing repulsive force, indicating an increase in contact area. This region is shown expanded in Figure 30, together with a fit (solid line) to Eq. 3, demonstrating that in this case there is good agreement with the plastic-deformation constriction resistance model. However, in general, the *quantitative* dependence of the contact resistance with applied force does not uniformly match the predictions of the model. Even in the post-ozonation samples, where contact resistances were relatively consistent, least-square power law fits to force-resistance curves demonstrate an average exponent of -0.19 with a standard deviation of 0.16 over 126 measurements. The average 'R-squared' value for these fits is 89%±10%, where R squared is the fraction of the total squared error for the power law model defined such that an R-squared value of 100% indicates a perfect fit. This variability is due to the inhomogeneity and compliant nature of the contamination layer, which is limiting the contact resistance through the variations in both topography and in chemical composition. The layer often maintains its electrical conductivity even as it deforms while the tip withdraws at significant distances from the gold surface – more than 5 nm in Figure 25, and more than 10 nm in Figure 27a. The initial electrical properties of the contamination layer, and also the distance to which it follows the withdrawing tip, vary from place to place on the surface, and even vary between contact events in nominally the same location, again due to differences in surface topography and because the layer itself evolves with each contact event. Based on the dramatic differences between the contact resistance of the layer in the pre- and post- ozonated surfaces, however, it is clear that the chemistry of the layer plays a significant role in determining its electrical properties. Observed changes may result from ozone exposure alone, but it is also possible that surface chemistry changed as a result of exposure to trace contaminants in the ozone cleaning system. Mechanically and with XPS, no strong evidence emerged which suggested the average layer thickness was appreciably decreased following exposure to ozone.

Summary

Interfacial force microscopy has been used to measure contact force and resistance between a micron-sized tip and an electroplated gold surface in a nitrogen ambient at room temperature, approximating single-asperity contact events in microsystem relays. Results demonstrate that contact resistance and adhesion are dominated by an adsorbed contamination layer which can be broken down with applied force and voltage. XPS and SIMS analysis show the layer to be composed mostly of hydrocarbons, with oxygen, nitrogen, carbon, and trace amounts of sulfur and iodine observed. The layer becomes dramatically more conductive following exposure to ozone, but no large changes in layer surface composition or thickness are observed. Finally, comparison with similar contact experiments performed in UHV show that post-ozone exposure contact measurements are representative of minimum contact resistance achievable with a single-asperity contact, and that adhesion forces are greatly diminished in the presence of the adsorbed contamination layer. It is clear from these results that, at least for Au/Au contacts, the contamination layer plays a very important role in passivating the strong metal/metal adhesive interaction. However, the evolution of the contamination layer under repeated contacts at high contact currents leading to eventual contact failure remains a major concern. This concern will only be diminished by further studies of the kind outlined here for various contact material combinations and, perhaps, various forms of contact coatings, e.g., graphitic or diamond-like carbons films.

5.0 Nanoscale Wiring Distribution

This chapter explores the concept of random interconnect slabs as an interposer between two other grids of dissimilar scale. The coarser grid represents a ``lithographically-accessible'' scale, while the finer grid represents a surface of molecular devices.

Nanoscale

The difference in physical scale between logic structures in molecular and traditional VLSI architectures creates a problem in signal exchange between the two. Figure 31 illustrates the concept in signal exchange through the notion of signal *launch planes*. In this context, signal launch planes represent grids of conductors that serve as connection points to another system, arbitrarily ordered but otherwised constrained in pitch and number by physical construction. To interface molecular circuits within a hybrid architecture to VLSI circuits, two such launch planes are required. Figure 31a illustrates a VLSI signal launch plane in which an arbitrary distribution signals are mapped to individual square conductors. The conductor center-to-center pitch is limited by lithography and fabrication technology, and in this example a representative pitch of 250~nm is chosen. The ``molecular launch plane'' Figure 31b is superimposed onto the VLSI launch plane as a grid having a much finer center-to-center pitch, in this example 10:1 or 25 nm. Dense molecular logic circuits require complex distributions of signals from the coarser VLSI grid to the finer molecular grid, as suggested by the example ``rat's nest'' wiring mapping in Figure 31c. The formation of this``rat's nest'' embodies the central problem in nano-interconnect distribution. Each of the molecular architecture schemes described so far, including those based on CPA, QCA, nanocell, and RCA approaches assume that input and output signals can be located as required, without discussion or justification that such flexibility can be assumed. Without an approach for redistribution, the most powerful concepts in molecular architecture described so far can be dramatically reduced in effectiveness due to *interconnect starvation*. Interconnect starvation is analogous to the situation of interconnection capacity-limited VLSI [102] in which circuit complexity is limited not by gate density but by the amount of wiring available. Hence, without an effective nano-interconnect distribution approach, molecular electronics are similarly limited in the complexity of circuits that they could otherwise express.

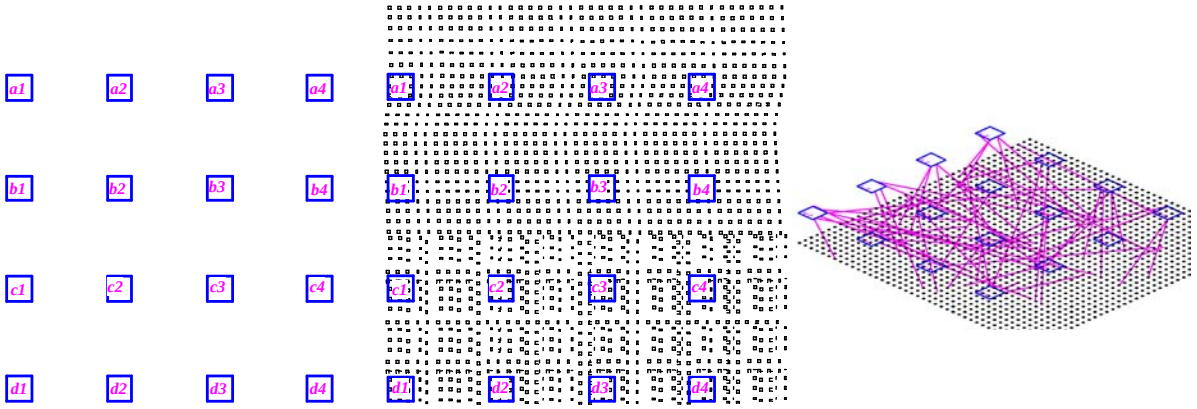


Figure 31. "VLSI-to-molecule" interconnection distribution. (a) Representation of 250~nm VLSI signal launch plane. (b) Superimposed 25~nm molecular signal launch plane. (c) Example distribution pattern.

Schemes for such nano-interconnect redistribution are inaccessible to traditional lithography due to scale. Even serial-write methods, such as e-beam lithography, that are capable of reaching these scales, have been demonstrated

only for simple, planar patterns, and no effective methods of multi-level e-beam have been reported. Shadow-masking approaches that can produce non-planar features are capable of producing only very simple features, since they rely on the deposition line-of-sight for feature generation.

Nanowires may offer a fundamentally different scheme for nano-interconnect. One possible scheme is suggested in Figure 32 This scheme exploits a number of randomly-oriented, dielectrically-sheathed nanowires embedded in a slab between two conducting grids. The nanowires in this approach must have a small diameter, smaller in dimension than the smallest average conductor grid pad spacing to prevent shorting conductors together. It is also necessary that the nanowires be insulated. It is well-known in statistical mechanics that conducting particles and sticks form a shorted network above a certain percolation threshold [103]. Figure 32 depicts two basic cases for nanowire distribution. In the first case (Figure 32a), the nanowires are vertically-directed, and ordered as a lattice. It is clear that such an arrangement provides no essential advantage, and from a connectivity standpoint, it is equivalent to directly bonding the two grid launch planes together. However, the second case (Figure32b) depicts a disordered arrangement of nanowires that produces a different connectivity pattern between the launch planes.

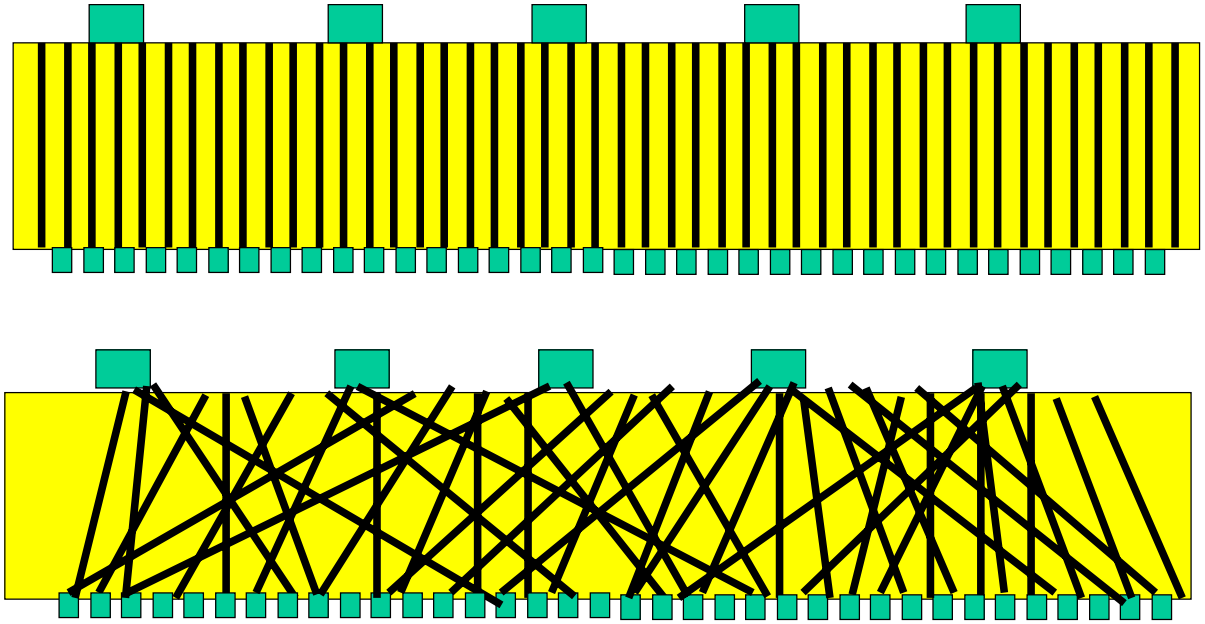


Figure 32. Conceptual nano-interconnection redistribution based on randomly-oriented nanowire slab. (top) Slab with vertically-directed, non-randomly oriented nanowire array. (bottom) Effect of randomized angles of nanowires on redistribution.

So while it is intuitively possible to argue that such a scheme (Figure 32b) might be better than "nothing" (Figure 32a) at producing a complex connective relationship between grids, it is necessary to consider two basic questions:

- How could such a grid be practically formed?
- Would any redistribution have useful properties as an interconnect redistribution network?

Addressing the first question requires definition of a possible fabrication scheme, that is made clear using an example of a silicon nanowire bundle shown in Figure 33. This figure illustrates a randomly ordered bundle of silicon nanowires formed using a proprietary process. Degenerately doped silicon is strongly conductive and upon thermal oxidation inherits a dielectric surface. The diameter of the silicon wire is further reduced through its consumption during oxidation. It is possible to stablilize a closely packed bundle through the infiltration of polymers (for example). This bundle can then be cleaved as suggested by the dashed horizontal lines to form a planar slab that can be interfaced to the VLSI and molecular launch plane grids. In this manner, it is possible to realize the Figure 32 construction. The statistical nature of the randomized interconnect distribution structure has the further practical advantage of tolerance to misalignment during assembly.

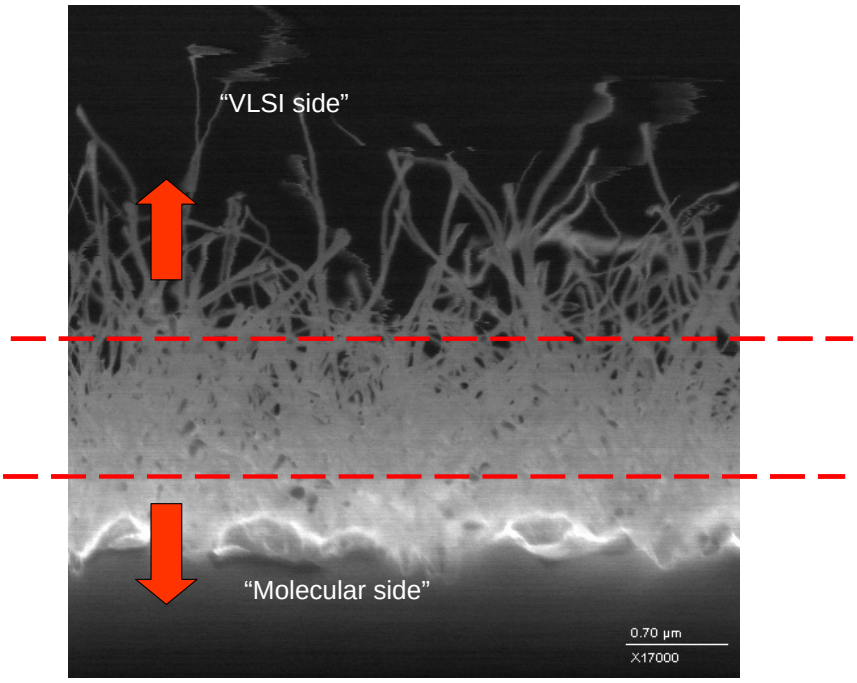


Figure 33. Potential mechanism for producing randomized wire distributed based on silicon microwires (courtesy of Saleem Zaidi, Gratings, Inc.)

Answering the second question requires developing some way of defining useful reshufflings and/or their properties and evaluating prospective distribution schemes against these expectations. Clearly, to be useful, a distribution must provide an effective spreading of signals into overlapping regions, so that a larger percentage of conductor locations in the fine grid becomes less ``starved'' in terms of signal diversity. For redistribution schemes displaying this characteristic, more analysis can provide insight into how the properties of these redistributions correlate with real-world redistribution networks, such as those in VLSI circuits. To understand the potential viability of the Figure 32b grid configuration, a simple, but more formalized model was established and series of simulations were conducted using this simple model with parameters as defined in the longitudinal cross section shown in Figure 34. The model uses three grids, two for the conductors and one for the nanowires. The conductor grids represent the fine (molecular) and coarse (VLSI) conductor grid launch planes offset in the z-axis at $z=0$ and $z=T$, respectively. Each launch is characterized as a square pad of infinitesimal thickness, having a dimension of one-half the x-y center-to-center conductor pitch, which is defined as p_F for the fine grid and p_C for the coarse grid. The third grid defines a set of nanowire conductors, that form a vertically directed having center-to-center pitch p_w and infinite spatial extent. Each wire is given a random angle $theta_{i,j}$, chosen from a Gaussian distribution about the vertical. The method for generating these angles, shown in Figure 35, employs a uniform distribution of lengths projected onto a plane offset slightly from the plane containing the pivot point of each angle ($z=T_{SLAB}$). Hence, the x-y position for a particular nanowire (i,j) at any z-point is unambiguous given its fixed pivot point ($x=i\ p_w, y=j\ p_w$). An angle is generated from the combination of a length chosen as a random, uniform distribution over $[0,r_{max}]$ and an angle chosen as a Gaussian distribution with negative z-axis-directed mean and standard deviation ϕ (a specified coning angle).

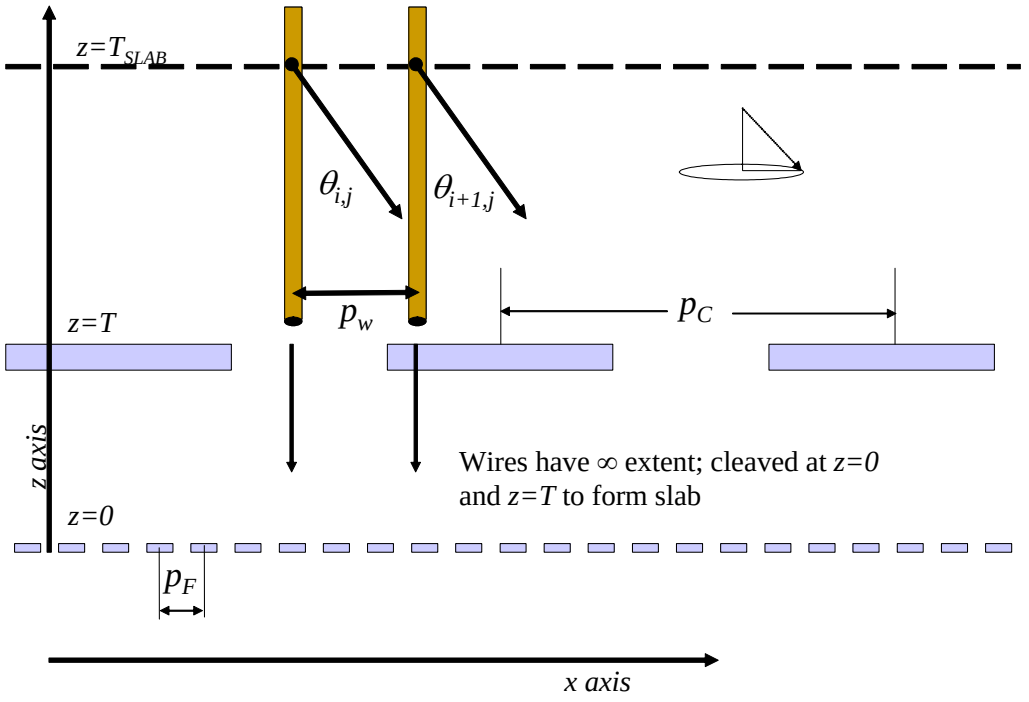


Figure 34. Parameters used to model random nanowire distribution network.

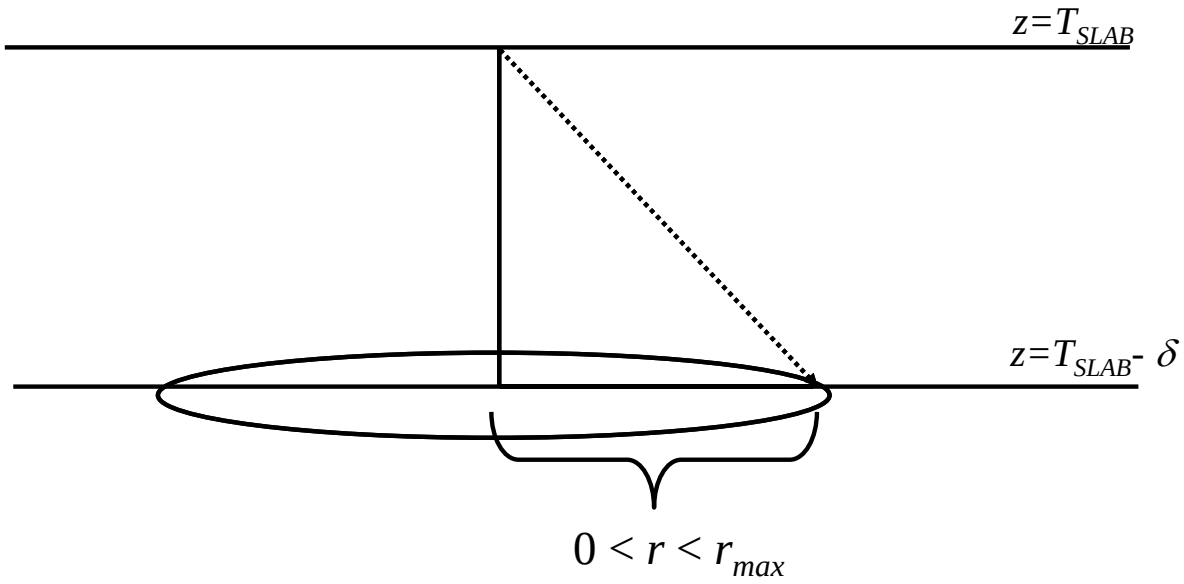


Figure 35. Method for generating random angles as projection onto offset plane.

The simulation code was based on several hundred lines of Mathematica with TCL support routines to generate 3-D (`.dxf` format) snapshots of the resulting physical structure, that could then be imported into Ashlar Vellum for visualization. The latter routines were useful in code debugging and provided a more visual confirmation that the modelled structures corresponded to the desired target configurations.

Connectivity for the purposes of these simulations was based on connections formed between particular nanowires, coarse pads, and fine grid regions. In these simulations, the fine grid did not correspond to a conductor grid {\em per se}, but rather a conductor "capture zone". By examining the number of different coarse pad connectiong to each fine grid zone through at least one nanowire, a rough measure of redistribution could be obtained through histograms and density plots of these quantities. In this case, the length of degree of a coarse pad (fine pad zone) is defined as the number of fine pad zones (coarse pads) intercepted through ``mediating" nanowires.

The results of initial simulations are shown in Figure 36 and Figure 37. The set of parameters assigned to the abstracted model were chosen to reflect a potential VLSI-to-molecular configuration of approximately correct dimensional order. In the Figure 36/Figure 37 case, p_w=25 nm, p_F=25 μm, p_C=250 nm, T=*1* μm, T_{SLAB}=4 μm, and the coning angle $\varphi=20^o$. Hence, the nanowire pitch is identical to fine pitch, and the coarse pitch is 10:1 coarser than either. In the simulation, the number of nanowires per spatial dimension were specified ($n_{wires}=120^2$ in this case), and statistics were gathered based on the resulting intersection geometries of individual nanowires. In the Figure 36 example, coarse pad degree distribution (Figure 36a) demonstrates a definite spread as a result of the random wire distribution (though no easily recognizable form). Fine grid zone distribution (Figure 36b reflects that most fine pads intercept only one coarse pad, though a rapidly declining number of fine pads intercept two and even three coarse pads. A spatial plot of coarse grid degree distribution (Figure 36c) reflects a spatial symmetry in degree tapering from the center. The partial surface plot (Figure 36d) of fine grid zone degree reflects a jagged distribution, implying that many fine pads do not intercept any coarse pads, and those that do intercept coarse pads also do not follow any easily recognizable spatial pattern. Based on the results of the initial simulation, 5,629 nets were generated involving 673 unique coarse pads and 5,629 fine pad zones. Physical plots run on this case are shown in Figure 38, that visually reflect a close correspondence to the simulated model. These plots visually suggest that significant signal redistribution is occuring. Another simple measure of redistribution is the count of nanowires intercepting fine pad zones that are not underneath coarse pads (n_{OFFSET}), since those pads would not be reachable without redistribution. In this simulation, 3,863 of the nanowires (26.8%) resulted in connections of this type.

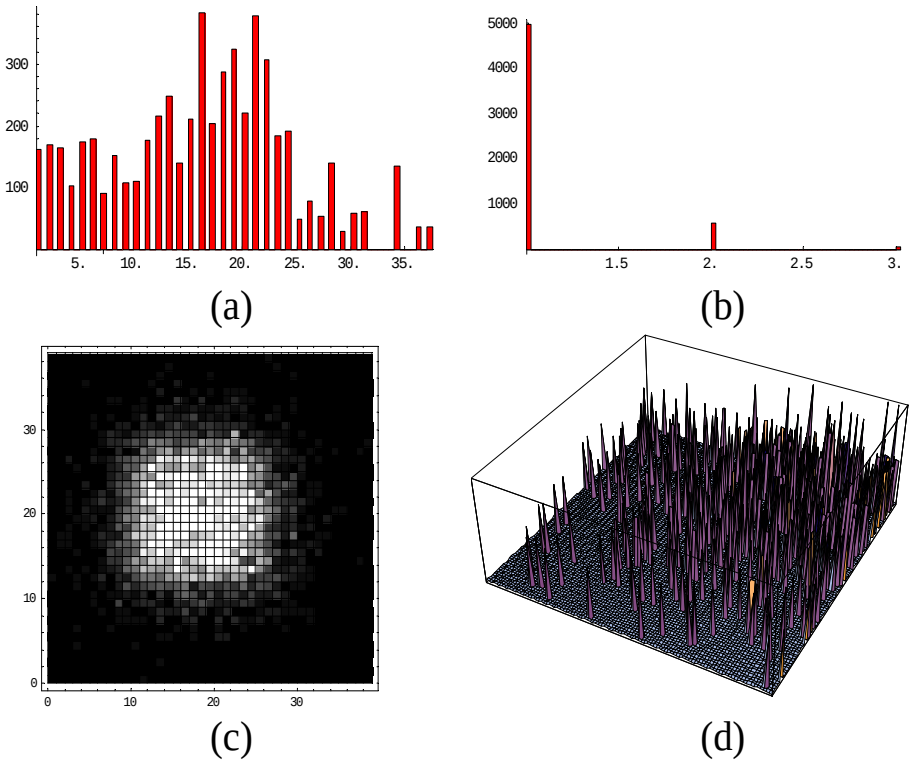


Figure 36. Results of initial nanowire distribution simulation with randomization on wire angles (p_w=25 nm, p_F=25 µm, p_C=250 nm, T=1 µm, T_{SLAB} =4 µm, φ=20°, n_{wires} = 14,400). Number of nets generated:5,629, with 673/1,330 unique coarse pads and 5,629 unique fine pad zones. 3,863 nanowires were mapped to fine pads not directly under coarse grid pads.

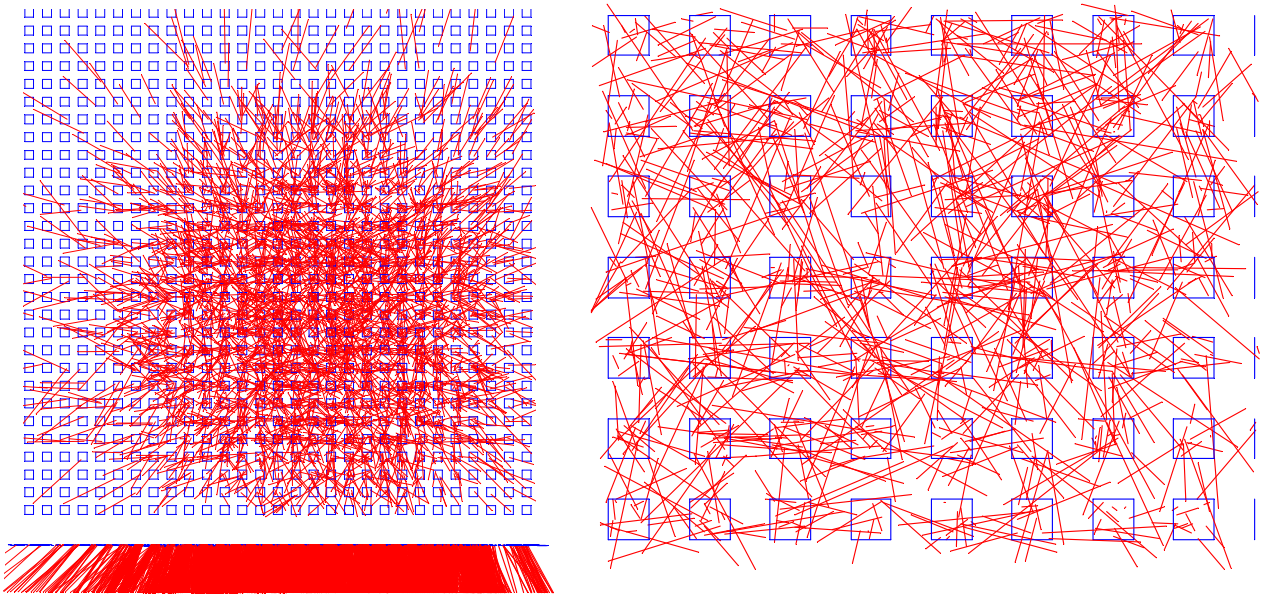


Figure 37. Physical appearance of particular simulation of randomized nano-wire distribution.

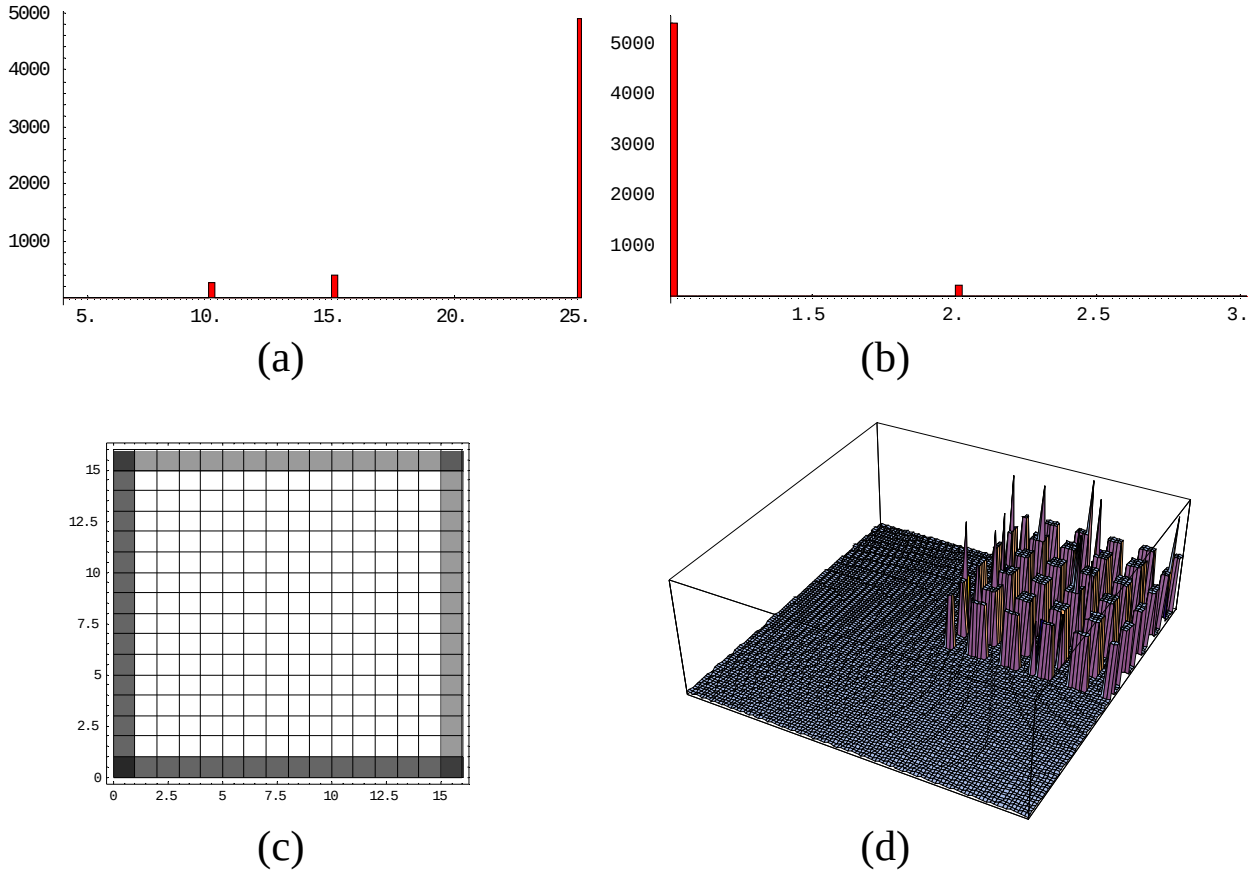


Figure 38. Nanowire distribution simulation without randomization (p_w=25 μm, p_F=25 nm, p_C=250 nm,T = 1 μm, T_{SLAB} = 4 μm, φ=0.05 , n_{wires}=14,400). Number of nets generated: 5,625. No nanowires were mapped to finepads unless directly under coarse grid pads.

How do these findings compare to the case where nanowires are not employed? This result can be also be simulated and compared to the obviously expected results. The initial simulation results were repeated using a very small coning angle (0.05 degrees (Figure 38)). The coarse pads in this example cover 25% of the fine pad zones. The coarse degree distribution (Figure 38a) should correspond to number of fine pad zones necessary to cover 25% of the area displaced by an *x-y* unit pitch of a coarse pad, which in this case is 25 fine pad zones. The histogram in (Figure 38a) reflects a small tail due to the cut off in the number of nanowires used in the simulation being less than integral multiple of this number. The fine pad zone degree (Figure 38b) should be degenerate (one coarse pad per fine pad). It is not, however, simply because a finite (non-zero) angle was used producing the small tail in the histogram. The coarse (Figure 38c) and fine pad zone (Figure 38d) spatial degree density plots reveal the definite, expected structure. Other statistics from the non-random analysis, though obvious, raise questions on the previous assertion that the number of nets indicate redistribution. Here, the number of nets is 5,625, nearly identical to the random case. The number of connected pad zones not directly under coarse pads n_{OFFSET} is of course identically zero, but this quantity alone may not be a sufficiently compelling indication that randomized redistribution is worth pursuing.

Distribution of Wirelength in Randomized Nanowire Networks.

The wirelength distribution for networks constructed in the manner of Figure 34 was examined under variations of angle, nanowire quantity, and slab thickness and distance from the origin plane. Prediction of wirelength, aided by Rent's rule analysis, was extensively studied in real-world networks [105], and wirelength of VLSI networks has often shown to follow a scale-free (power law) distribution. Davis *et al.* in particular developed methods for fitting data from real VLSI designs (such as shown in Figure 39) to such expressions [105].

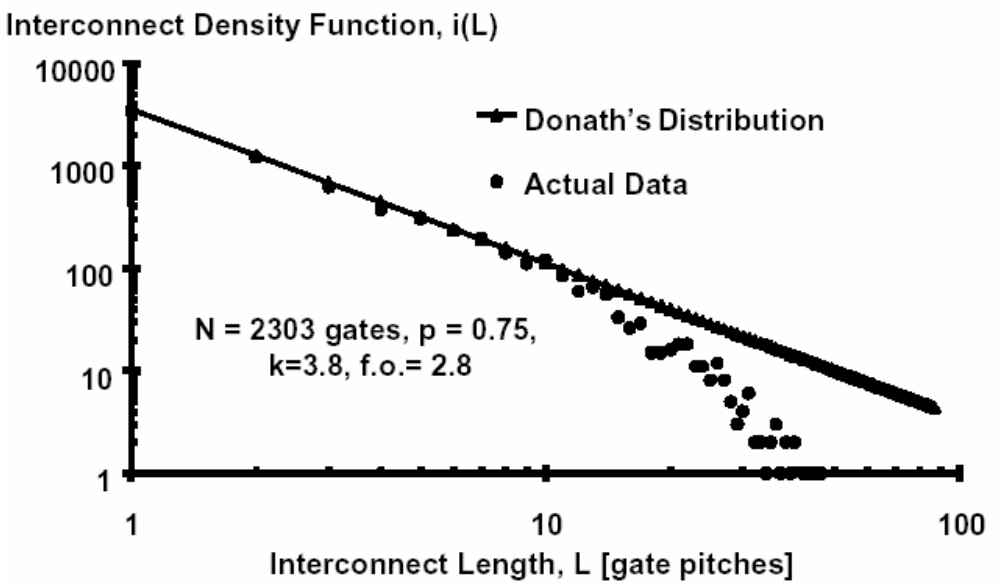


Figure 39. Empirical fitting of VLSI data to scale-free (power law) relationship [108].

It is somewhat surprising to find that the randomized nanowire networks also follow a scale-free distribution, as evidenced in Figure 40. In these examples, the number of nanowires per axis was increased from 20 to 400. To produce these distribution plots, it was necessary to scale and truncate the raw wirelength numbers, so that discretized bins are produced. When the number of connected nanowires is increased to the point of statistical significance (for example, at 20 nanowires per axis, the form of distribution is not well-defined), the emergent form of the distribution is striking. Not only do the basic shapes of the distributions in Figure 40 agree with the example shown in Figure 39, but the characteristic spreading of the statistical data at large wirelengths also appears in both cases. Hence, using a random growth model fundamentally different from Barabasi *et al.*[106] (and more recent attempt to apply their work to interconnect [107]), these results suggest a new approach for generating scale-free models whose properties mimic the behavior of human-mediated complex VLSI networks.

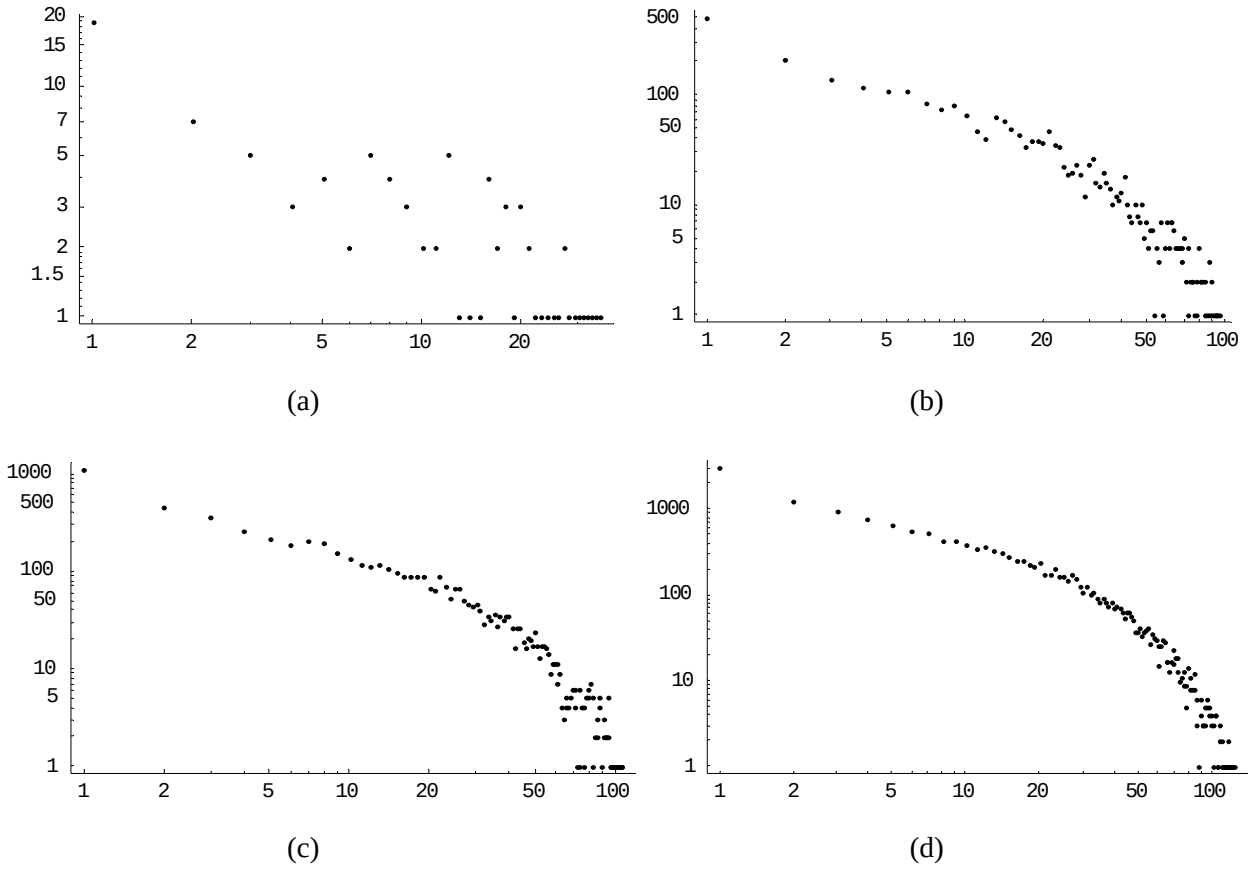


Figure 40. Scale-free distributions produced from nanowire networks. In these plots, *x*-axis is wirelength, and the *y*-axis corresponds to the number of wires at particular wirelengths. (a) 20x20 nanowires. (b) 100x100 nanowires. (c) 200x200 nanowires. (d) 400x400 nanowires. (p_W=p_F= 25 nm, φ=30°, p_C= 250 nm, T_{SLAB} = 4 μm, T = 1μ.)

The effects of angle on wirelength distribution using the Figure 37 example as baseline are shown in Figure 41 and Figure 42. The scale-free characteristic of the distributions are recognizable, even at low coning angles. As coning angle increases for a fixed number of nanowires, the spread in distribution of nanowires increases along with a corresponding reduction in magnitude. Spread in the distribution at longer wirelengths becomes gradually more pronounced at higher angles. At coning angles above 50° (Figure 42), the spreading becomes significant enough to distort the distribution. At pronounced angles (above 75°) the distribution is essentially unrecognizeable.

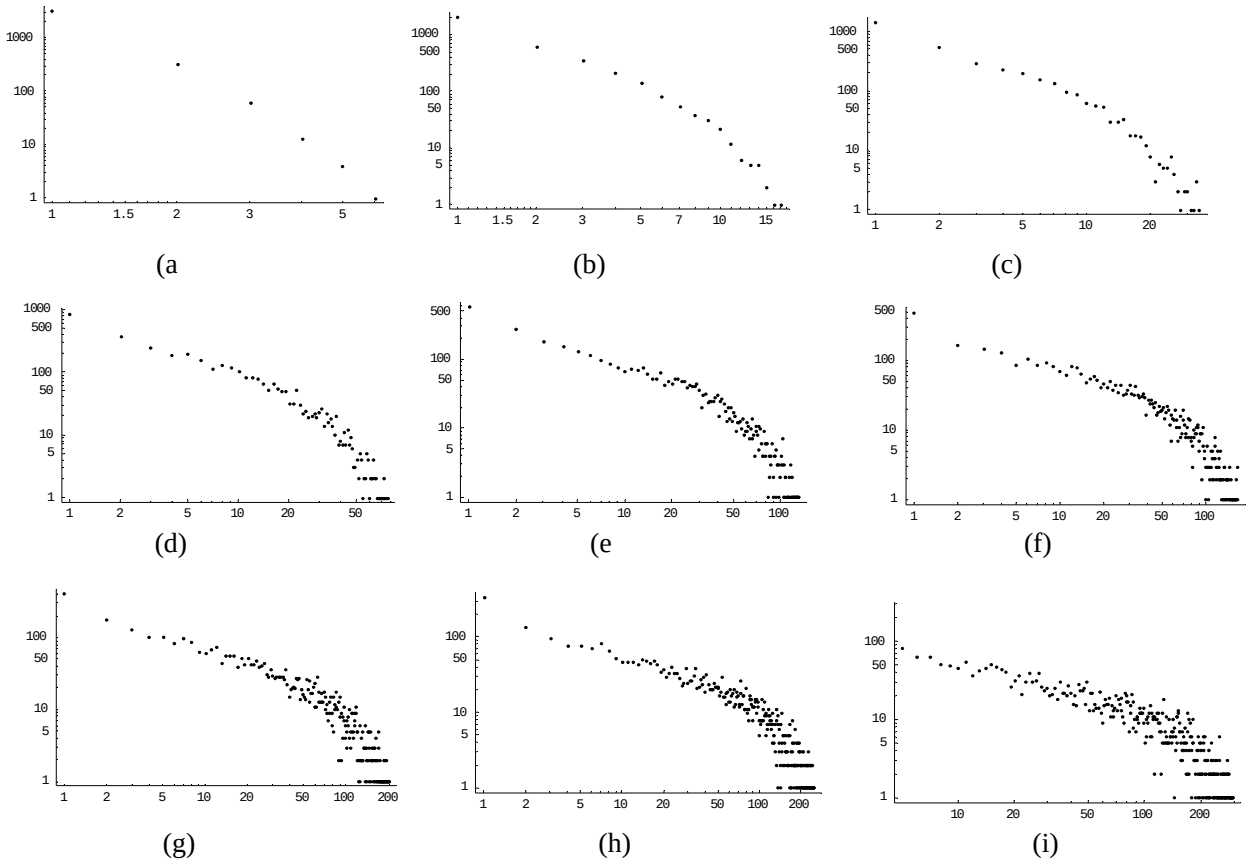


Figure 41. Effects of coning angle on wirelength distribution. (a) φ = 5°. (b) φ = 10°. (c) φ = 15°. (d) φ = 20°. (e) φ = 25°. (f) φ = 35°. (g) φ = 40°. (h) φ = 45°. (i) φ = 50°. ($p_W = p_F = 25$ nm, $p_C = 250$ nm, $T_{SLAB} = 4$ μm, $T = 1$ μm, $n_{WIRES} = 14{,}400$.

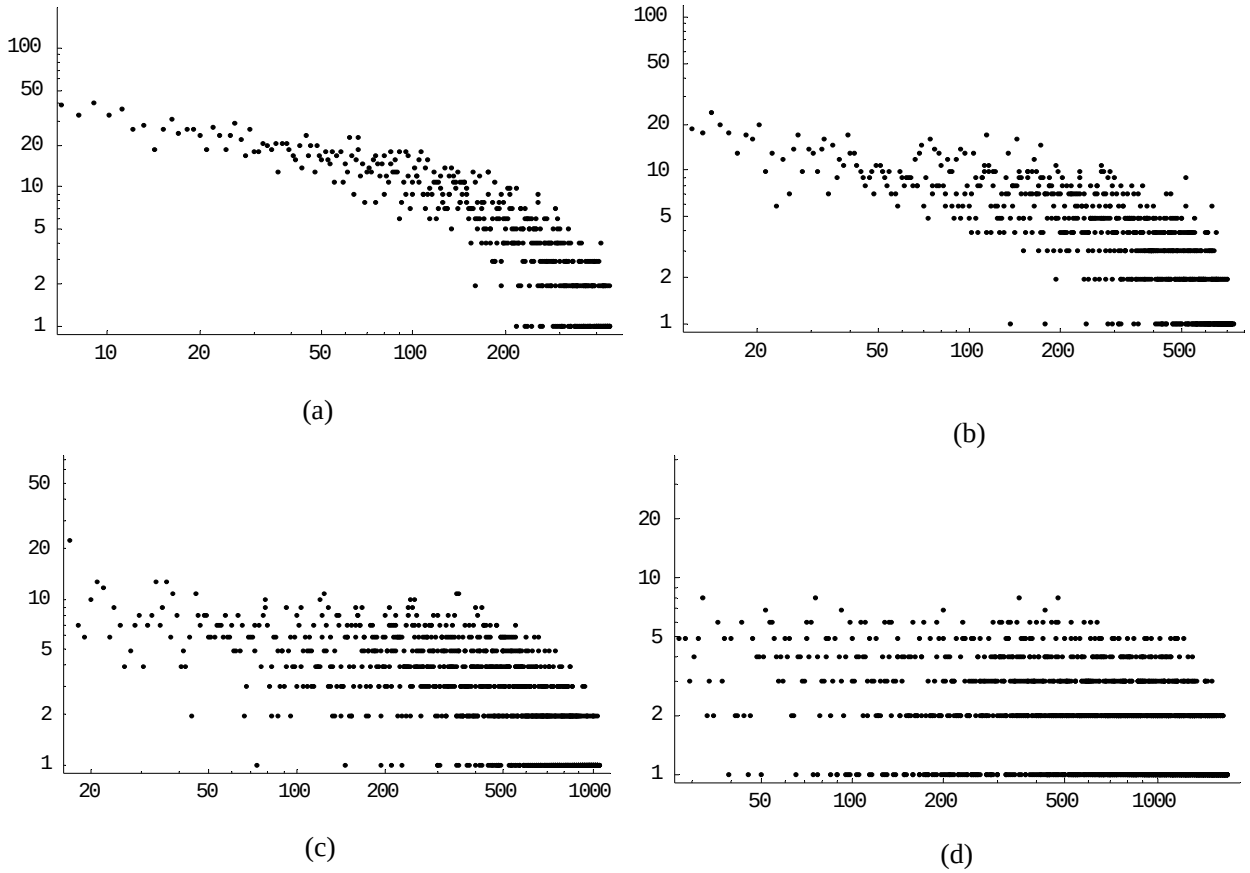


Figure 42. Effects of coning angle on wirelength distribution (concluded). (a) φ = 65°. (b) φ = 70°. (c) φ = 80°. (d) φ = 85°. (p_w = p_F= 25 nm, p_C= 250 nm, T_{SLAB} = = 4 μm, T = 1 μm, n_{WIRES} =14,400.

The variation in distribution due to changes in the slab thickness (T_{SLAB}) and final thickness (T) were also examined. Typical results are presented in Figure 43 where final thickness of a typical configuration is varied from 0.1 μm to 3 μm, holding T_{SLAB} and angle (φ) unchanged. Results demonstrating the effects of T_{SLAB} variation with constant T (except below T_{SLAB} = 1 μm) are shown in Figure 44. The results for the plots in Figure 44 are of practical interest since in fabrication it is likely that the final thickness must be held to more precise values for overall process control. Simulations in which the ratio of starting and final thickness are held constant were also performed. Plots showing the effect of thickness variation under the condition $T = T_{SLAB}$ are shown in Figure 45.

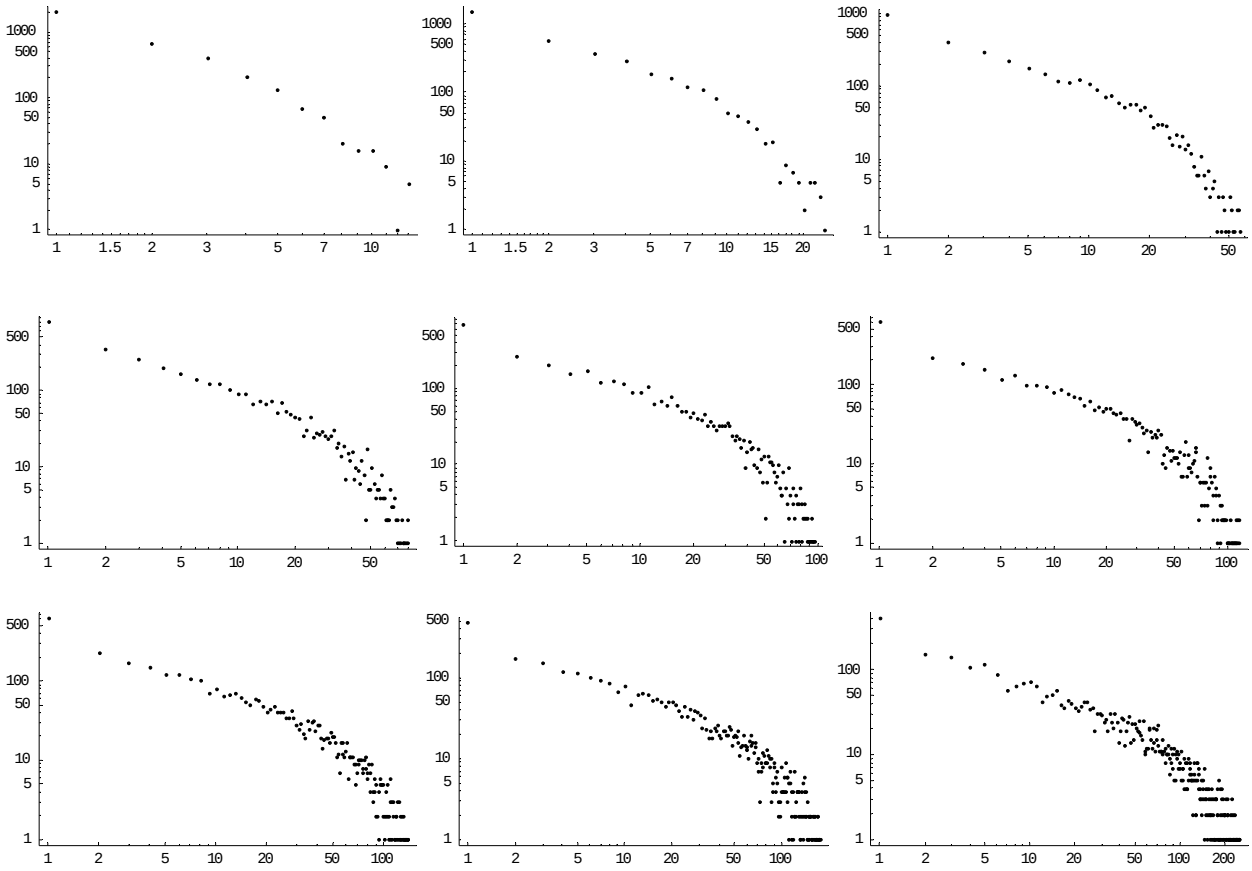


Figure 43. Effects of final thickness (*T*) on wirelength distribution. (a) *T* = 0.1μm. (b) *T* = 0.2μm. (c) *T* = 0.5μm. (d) *T* = 0.75μm. (e) *T* = 1.0μm. (f) *T* = 1.25 μm. (g) *T* = 1.5μm. (h) *T* = 2μm. (i) *T*=3μm. (p_w=p_F= 25 nm, p_C = 250 nm, T_{SLAB} =4 μ, φ=30°, n_{WIRES} = 14,400.)

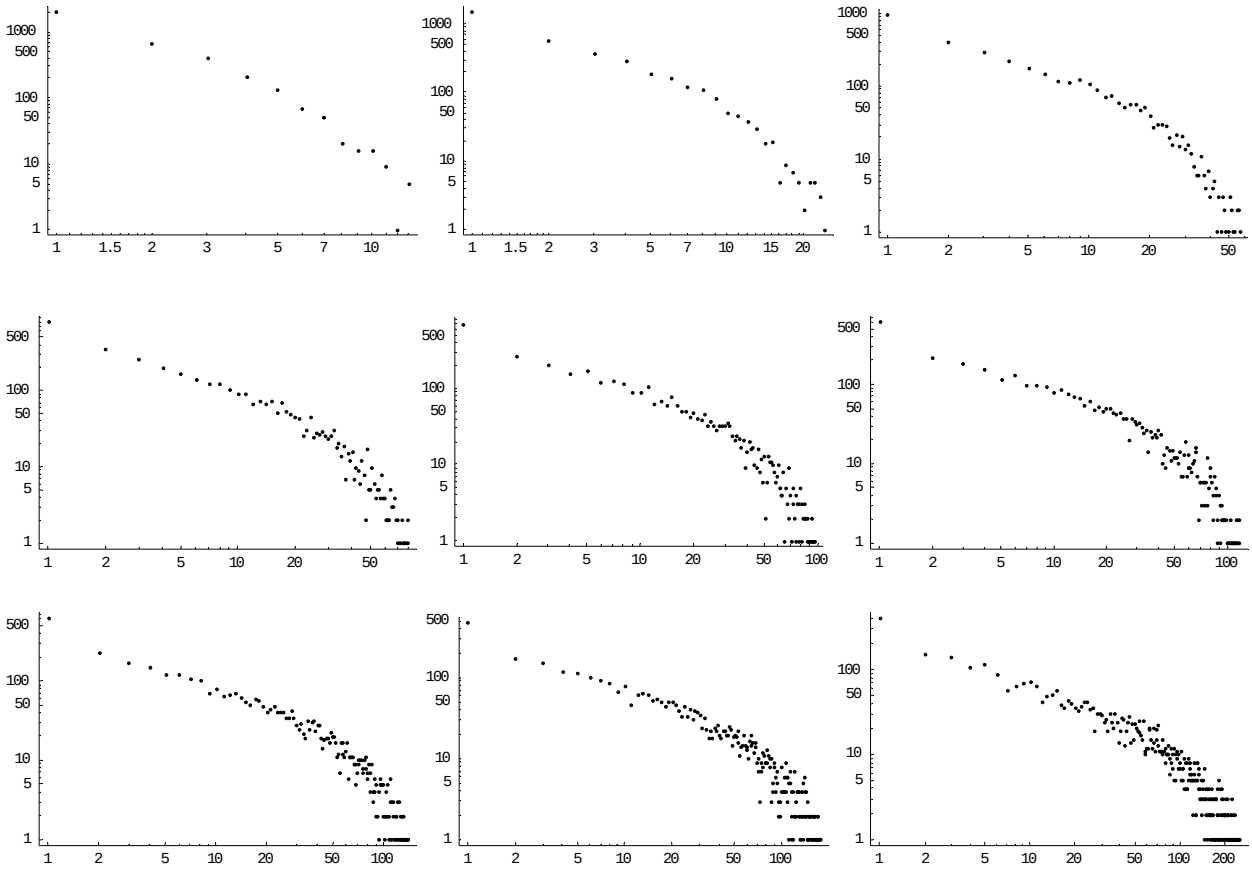


Figure 44. s316 Effects of slab thickness T_{SLAB} on wirelength distribution. (a) T_{SLAB} = 0.25 μm, T = 0.1 μm. (b) T_{SLAB} =0.5 μm, T = 0.2 μm. (c) T_{SLAB} = 1.0 μm, T = 0.5 μm. (d) T_{SLAB} = 2.0 μm, T = 1.0 μm. (e) T_{SLAB} =4.0 μm, T = 1.0 μm. (f) T_{SLAB} =6.0 μm, T = 1.0 μm. (g) T_{SLAB} =8.0 μm, T = 1.0 μm. (h) T_{SLAB} = 16.0 μm, T = 1.0 μm. (i) T_{SLAB} =32.0 μm, T = 1.0 μm. (p_w=p_F= 25 nm, p_C = 250 nm, T_{SLAB} =4 μ, φ=30°, n_{WIRES} = 14,400.)

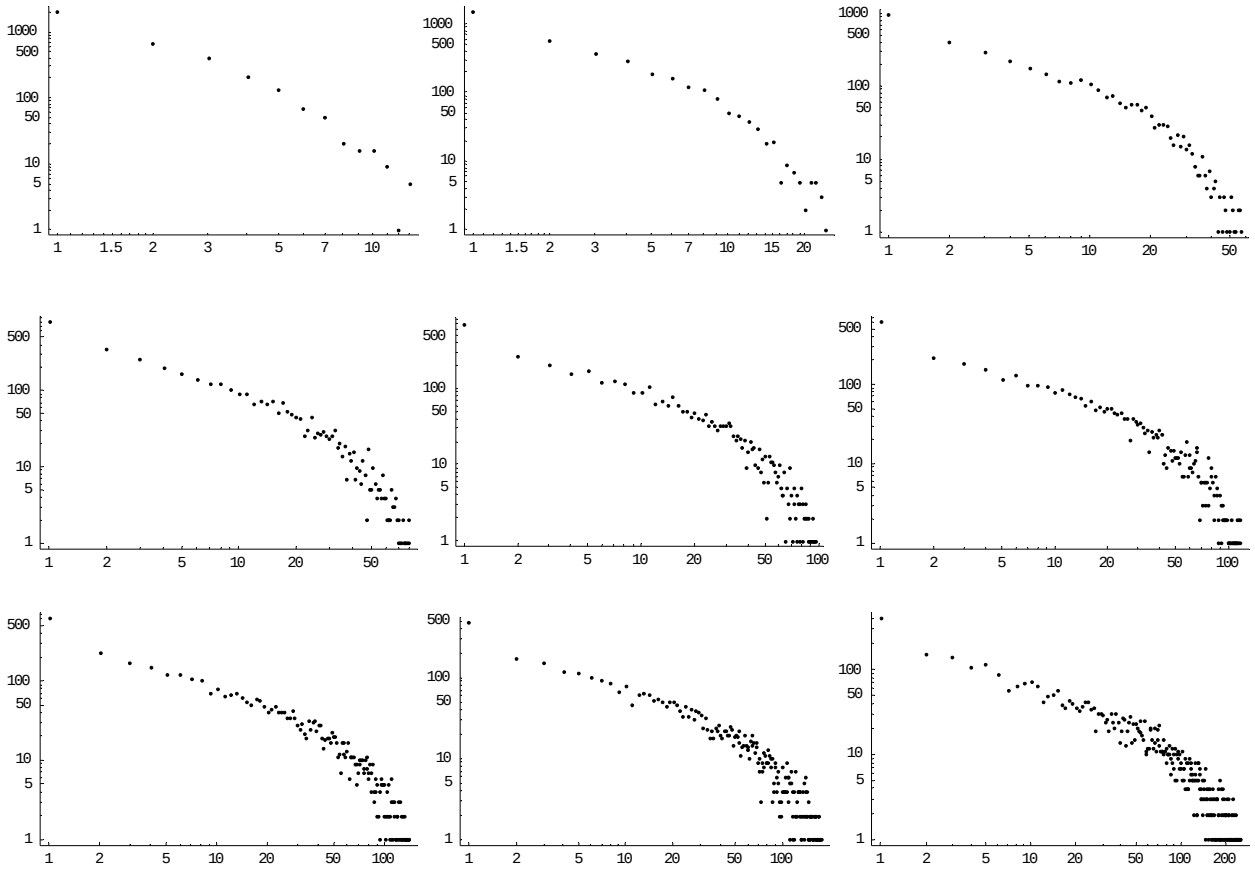


Figure 45. Effects of variation of thickness with constant (T_{SLAB} / T). (a) 250 nm. (b) 2 μm. (c) 4 μm. (d) 8 μm. (p_w=p_F= 25 nm, p_C = 250 nm, T_{SLAB} =4 μ, φ=30°, n_{WIRES} = 14,400.)

These simulations reveal that the effects of thickness variations on wirelength distribution are in a gross sense similar to the effects of advancing angle. In fact, the emergence of scale-free behavior seems quite robust, and this behavior disappears only in cases that approach nearly pathological extremes (e.g., coning angles of ~0° or π/4, or T_{SLAB}= T and $T >> p_C$). In "reasonable" cases, where the coning angle is in the range $20^o < \varphi < 45^o$ and the $p_C << T < 0.25\ T_{SLAB} < 100\ p_C$, a scale-free behavior is produced in grids of nanowires subject to the conditions discussed in this chapter.

6.0 Conclusions

Reconfigurability has become very important to 21st century systems. This report addressed reconfigurable systems with an emphasis on wires. Wires are important as they related to the interconnections between components. Take a collection of components and throw them on a table, and you have a bunch of components. Connect the same components together, and you have architecture. Being able to recompose the same components into different architectures under software control gives you a reconfigurable architecture. The most well known real-world example is the field programmable gate array (FPGA). It represents a sort of "digital parts cabinet" in which a user, through a software program, can define the connective relationship between logic and memory to form a digital system. It is partly an illusion, because the physical structure is pre-defined, and the FPGA is programmed only by re-arranging charge configurations in this pre-built piece of silicon. It is in fact necessary to pre-build a superset of wiring possibilities, a subset of which is selected for any one user design. If an illusion (or more correctly an emulation) the illusion is becoming more effective as driven by Moore's law. At the time of this writing, systems with more than 10 million effective gates can be defined this way.

In this report, extensions of the software-definability concepts that give rise to reconfigurable behavior were examined. The ideas of reconfigurable systems were outlined, from which a taxonomy was defined. Within this taxonomy are found the elements of digital, analog, microwave, power, pathways, mechanisms, and matter. In the limit, the surface and bulk properties of matter itself would be in effect software-definable. Whether or not programmable matter is possible, and we may not know the answer even in this century, the ideas of a discipline for reconfigurable systems emerge. At the first level, we consider where the "knobs" or degrees of freedom are or can be introduced. At the second level, we establish how those knobs are adjusted, offline or real-time, remotely or *in situ*. Computer science provides concepts of computability and decidability, the whole of algorithms, resolvable ultimately (if a problem can be computed at all) into an elemental set of 0-1 programming decisions. These in turn drive the "knobs", and reconfigurable systems becomes the framework under which algorithms operate upon the "real world" as those programming decisions are impressed into the knobs of a particular system.

This document has reviewed a number of those key advances in reconfigurable systems technology. The field is still emergent, and this report, if updated in about five years, would undoubtedly cover in more depth the advances in reconfigurable analog and power systems. Thematically, the emphasis of this report has been on "smart wires" and technologies relating to making reconfigurable pathways. Pathways are essential in this embryonic field, since they define the connective relationships between components, without which architectures are reduce to a mere collection of loose components.

References

[1] DeHon,A. "Reconfigurable Architectures for General-Purpose Computing", Massachusetts Institute of Technology Artificial Intelligence Laboratory, *A.I. Technical Report No. 1586*, October 1996.
[2] Katz, R.; Swift, G.; Shaw, D. "Total dose responses of Actel 1020B and 1280A field programmable gate arrays (FPGAs)", Third European Conference on Radiation and its Effects on Components and Systems, 1995.
[3] Penrose, N.B.; Gibson, D.A.; McMillian, G.B. "GaAs FPGA for Space", Phillips Laboratory Technical Report, *DTIC Report ADB242240*, February 1996.
[4] "Honeywell and Atmel Team to Develop Radiation-Hardened Configurable FPGAs", press release, Honeywell Corporation, 24 July 1998 (http://www.ssec.honeywell.com/aerospace/new/072498.html).
[5] McCabe, J. "Radiation Hardened Field Programmable Gate Array", *Proceedings of the 1998 Military And Aerospace Applications Of Programmable Devices And Technologies (MAPLD) Conference*, Johns Hopkins University (Laurel MD), September 1998, (proceedings available at http://klabs.org).
[6] Specifications for the 4K and 64K Orion FPGAs are available at the Mission Research Corporation Microelectronics Center website (http://www.mrcmicroe.com/).
[7] Mavis, D.; Cox, B.; Adams, D.; Green, R. "A Reconfigurable, Non-Volatile, Radiation-Hardened Field Programmable Gate Array (FPGA) for Space Applications", *Proceedings of the 1998 Military And Aerospace Applications Of Programmable Devices And Technologies (MAPLD) Conference e*, Johns Hopkins University (Laurel MD), September 1998, (proceedings available at http://klabs.org).
[8] Rockett, L.R. "A design based on proven concepts of an SEU-immune CMOS configuration data cell for reprogrammable FPGAs", *Microelectronics Journal*, **32**(2):99-111, February 2001.
[9] Anghel, L.; Alexandrescu, D.; Nicolaidis, M. "Evaluation of a soft error tolerance technique based on time and/or space redundancy" *Proceedings of 13th Symposium on Integrated Circuits and Systems Design*, 2000, pp. 237 -242
[10] Wang, J.J.; Katz, R.B.; Sun, J.S.; Cronquist, B.E.; McCollum, J.L.; Speers, T.M.; Plants, W.C. "SRAM based re-programmable FPGA for space applications", IEEE Transactions on Nuclear Science, Part I, 46(6):1728-1735, December 1999.
[11] Fuller, E.; Caffrey, M; Blain, P.; Carmichael, C.; Kalsa, N.; Salazar, A. "Radiation Test Results of the Virtex FPGA and ZBT SRAM for Space-Based Reconfigurable Computing", *Proceedings of the 1998 Military And Aerospace Applications Of Programmable Devices And Technologies (MAPLD) Conference*, Johns Hopkins University (Laurel MD), September 1998, (proceedings available at http://klabs.org)
[12] Bergmann, N.W. and A.S. Dawood, "Reconfigurable Computers in Space: Problems, Solutions, and Future Directions", *Proceedings of the 1999 Military And Aerospace Applications Of Programmable Devices And Technologies (MAPLD) Conference*, Johns Hopkins University (Laurel MD), September 1999, (proceedings available at http://klabs.org).
[13] Brodrick, D.; Dawood, A.; Bergmann, N.; Wark, M. "Error detection for adaptive computing architectures in spacecraft applications" *Proceedings of the 6th Australasian Computer Systems Architecture Conference*, 2001, pp. 19-26.
[14] MacQueen, D.M.; Gingrich, D.M.; Buchanan, N.J.; Green, P.W. "Total ionizing dose effects in a SRAM-based FPGA", *Radiation Effects Data Workshop*, 1999, pp. 24 -29
[15] MacQueen, D.M. " ", Master's thesis, University of Alberta, CA (http://csr.phys.ualberta.ca/~macqueen/thesis/).
[16] Desplats, R.; Perdu, P.; Benteo, B.; Grangy, A. "Electron Beam Testing of FPGA Circuits", *Microelectronics and Reliability*, 39(6), pp. 963-968, June 1999.
[17] Ceschia, M.; Bellato, M. ; Menichelli, M., Papi, A.;Wyss,J.;Paccagnella, A. "Heavy ion irradiation of SRAM-based FPGA's", *Proceedings of the 2001 Military And Aerospace Applications Of Programmable Devices And Technologies (MAPLD) Conference, ,* Johns Hopkins University (Laurel MD), September 2001, (proceedings available at http://klabs.org).
[18] Li, Y.; Li, D.; Wang Z. "A new approach to detect-mitigate-correct radiation-induced faults for SRAM-based FPGAs in aerospace applications", *Proceedings of the IEEE 2000 National Aerospace and Electronics Conference*, pp. 588 -594.
[19] Alfke, P. and R. Padovani "Radiation Tolerance of High-Density FPGAs", *Proceedings of the 1998 Military And Aerospace Applications Of Programmable Devices And Technologies (MAPLD) Conference*, Johns Hopkins University (Laurel MD), September 1998, (proceedings available at http://klabs.org).
[20] IEEE Std 1451.2, Standard for a Smart Transducer Interface for Sensors and Actuators - Transducer to Microprocessor Communication Protocols and Transducer Electronic Data Sheet (TEDS) Formats.

[21] Thomke, Stefan H. "The role of flexibility in the development of new products: An empirical study", *Research Policy*, **26**(1):105-119, March 1997.
[22] Lyke, J.; Wilson, W.; Contino, P. "MEMS-Based Reconfigurable Manifold", *Proceedings of the 2000 Military And Aerospace Applications Of Programmable Devices And Technologies (MAPLD) Conference*, Johns Hopkins University (Laurel MD), September 2000, (proceedings available at http://klabs.org).
[23] Tyson, S.; Wicker, G.; Lowrey, T.; Hudgens, S.; Hunt, K. "Nonvolatile, high density, high performance phase-change memory", *Proceedings of the IEEE 2000 Aerospace Conference*, vol 5:385-390.
[24] Bernacki, S.; Hunt, K.; Tyson, S.; Hudgens, S.; Pashmakov, B.; Czubatyj, W. "Total dose radiation response and high temperature imprint characteristics of chalcogenide based RAM resistor elements", *IEEE Transactions on Nuclear Science*, Part 3, **47**(6):2528 -2533.
[25] "Digital Crosspoint Switches - Frequently Asked Questions", application not, *Doc #DCS_FAQ_TechNote*, Rev 2.0, July 2001, http://www.icube.com/.
[26] Jun Z.; Chang L. "Development of a novel micro electromechanical tunable capacitor with a high tuning range" *Conference Digest of the 58th Device Research Conference*, 2000, pp. 111 -112.
[27] Lubecke, V.M.; Barber, B.; Chan, E.; Lopez, D.; Gammel, P. "Self-assembling MEMS variable and fixed RF inductors", *2000 Asia-Pacific Microwave Conference*, pp. 201-204.
[28] "Using a DCS Device as an Analog Switch", application note, *Doc #IQX_AnalogSwitch_TechNote*, Rev 2.0, July 2001, http://www.icube.com/.
[29] Sackinger, E.; Lee, E.K.F.; Gulak, P.G. "A CMOS Field-programmable Analog Array". *IEEE Solid-State Circuits Conference. (ISSCC)*, , 13 February 1991, pp. 186 - 314.
[30] Gulak, P.G. "Field programmable analog arrays: past, present and future perspectives", *Proceedings of International Conference on Microelectronics*, 6 November 1995, pp. 123-126.
[31] Lee, E.K.F.; Gulak, P.G. "Field programmable analogue array based on MOSFET transconductors", *Electronics Letters* 2 January 1992, **28**(1): pp. 28 - 29.
[32] Schweber, B. "Programmable Analog ICs: Designer's Delight or Dilemma?", *Electronic Design News*, 13 April 2000, pp 72-84.
[33] Bindra, A. "Field Programmability Pervades Analog Devices", Electronic Design, 9 July 2001, pp. 56--62..
[34] Edwards, R.T.; Strohbehn, K.; Jaskulek, S.E. "A field-programmable mixed-signal array architecture using antifuse interconnects", *Proceedings from the 2000 IEEE International Symposium on Circuits and Systems*, Geneva, 2000, pp. 319 -322.
[35] Nguyen, C.T.-C.; Katehi, L.P.B.; Rebeiz, G.M. "Micromachined devices for wireless communications", Proceedings of the IEEE 86(8):1756-1768, Aug. 1998
[36] Nguyen, C.T.-C. "Frequency-selective MEMS for miniaturized low-power communication devices", *IEEE Transactions on Microwave Theory and Techniques*, 47(8):1486-1503, August 1999.
[37] Vig, J.R.; Yoonkee K. "Noise in microelectromechanical system resonators", *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 46(6):1558-1565, November 1999.
[38] David Vreeland, C.G. Christodoulou, James Lyke, "Application of Chalcogenide-Based Materials For Use in Programmable Circuitry for a Microstrip Antenna," *2001 IEEE AP-S International Symposium Digest USRI*, p. 374
[39] Lakkas, G.; Duduman, B. "The ACPI Advantage for Powering Future-Generation Computers", Electronic Design News, 20 September 2001, pp. 91-94.
[40] Comtois, J. "Structures and Techniques for Implementing and Packaging Complex, Large Scale Microelectromechanical Systems Using foundry Fabrication Processes", PhD dissertation, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, *AFIT/DS/ENG/96-04*, 1996.
[41] Mita, Y.; Kaiser A; Stefanelli, B.; Garda, P.; Milgram, M.;Fujita, H. "A Distributed microactuator Conveyance System with Integrated Controller", *Proccedings of IEEE SMS'99*, 1999, pp. 18-21.
[42] Toffoli, T. "Programmable Matter Methods", *Future Generation Computer Systems*, 16(1999):187-201.
[43] Y. Kinashi, R. Linderman, and J. Lyke, "DITP: A Flexible miniaturized Sensor Fusion Processor for next Generation Interceptor Seekers and Surveillance Sensors", *Proc. Of the 7th Annual AIAA/BMDO Technology Readiness Conference and Exhibit*, Colorado Springs, CO. August 3-6, 1998.
[44] Mitola, J. "The software radio architecture", *IEEE Communications Magazine* **33**(5):26-38
[45] *Open Systems Acquisition & Supportability Guide*, Next Generation Computer Resources (NGCR). *NGCR Document No. AST 003* ver. 1.0, published by Navy SPAWAR, 31 December 1996.
[46] Travis, B. "Smart Conditioners Rub Out Sensor Errors", *Electronic Design News*, 1 February 2001.
[47] Abbott, R.; Cannon S. "Adaptive Computer Systems", *Proceedings of the IEEE 2001 Aerospace and Electronics Conference.*

[48] Lyke, J; Avery; K.; Brezna, P. "BMDO/AFRL Partnership: Advances in Data Handling Systems for Space Experiment Control", *Proceedings of the 2000 Military And Aerospace Applications Of Programmable Devices And Technologies (MAPLD) Conference*, Johns Hopkins University (Laurel MD), September 2000, (proceedings available at http://klabs.org).
[49] Joint Test Action Group (JTAG), *IEEE Standard 1149.1*, 1990.
[50] Wolfram, S. *Cellular Automata and Complexity.* Addison-Wesley, New York (1994).
[51] Bursky, D. "High-Density FPGAs Take on System ASIC Features and Performance Levels", *Electronic Design*, 18 September 2000, pp. 97-113, 124.
[52] "Digital Crosspoint Switches - Frequently Asked Questions", application not, *Doc #DCS_FAQ_TechNote*, Rev 2.0, July 2001, http://www.icube.com/.
[53] "Using a DCS Device as an Analog Switch", application note, *Doc #IQX_AnalogSwitch_TechNote*, Rev 2.0, July 2001, http://www.icube.com/.
[54] Sackinger, E.; Lee, E.K.F.; Gulak, P.G. "A CMOS Field-programmable Analog Array". *IEEE Solid-State Circuits Conference. (ISSCC)*, , 13 February 1991, pp. 186 - 314.
[55] Gulak, P.G. "Field programmable analog arrays: past, present and future perspectives", *Proceedings of International Conference on Microelectronics*, 6 November 1995, pp. 123-126.
[56] Lee, E.K.F.; Gulak, P.G. "Field programmable analogue array based on MOSFET transconductors", *Electronics Letters* 2 January 1992, **28**(1): pp. 28 - 29.
[57] Schweber, B. "Programmable Analog ICs: Designer's Delight or Dilemma?", *Electronic Design News*, 13 April 2000, pp 72-84.
[58] Travis, B. "Smart Conditioners Rub Out Sensor Errors", *Electronic Design News*, 1 February 2001.
[59] Bindra, A. "Programmable SoC Delivers a New Level of System Flexibility", *Electronic Design*, 6 November 2000, pp. 74-78.
[60] Bindra, A. "Field Programmability Pervades Analog Devices", Electronic Design, 9 July 2001, pp. 56--62..
[61] Keymeulen, D.; Zebulum, R.S.; Jin, Y.; Stoica, A. "Fault-tolerant evolvable hardware using field-programmable transistor arrays", *IEEE Transactions on Reliability*, September 2000, **49**(3): pp. 305 - 316
[62] Meyerson, B. "SiGe Upgrades Wireless, Soft Radio", *Electronic Engineering Times*, 7 February 2000., p. 80.
[63] Lee, Y.-S. and A. C.-H. Wu, "A Performance and Routability -Driven Router for FPGA's Considering Path Delays", *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 16(2):179-185, February 1997.
[64] Alexander, M.J. and G. Robins, "New Performance-Driven FPGA Routing Algorithms", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, **15**(12):1505-1517, December 1996.
[65] Joint Test Action Group (JTAG), *IEEE Standard 1149.1*, 1990.
[66] Comtois, J. "Structures and Techniques for Implementing and Packaging Complex, Large Scale Microelectromechanical Systems Using foundry Fabrication Processes", PhD dissertation, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, *AFIT/DS/ENG/96-04*, 1996.
[67] Ragnar Holm, *Electric Contacts Handbook*, Springer-Verlag, 1958, p 17
[68] Ezekiel Kruglick, *Microrelay Design, Performance, and Systems*, dissertation, Berkeley Sensor and Actuator Center, University of California, Berkley, p 42
[69] Allan, R. "Lateral CMOS Process Yields 7.4V MOSFET With On-Resistance of 6 milli-ohms", *Electronic Design*, 12 July 1999, pp 36-38.
[70] J. Schimkat, *Contact Measurements Providing Basic Design Data for Microrelay Actuators*, Sensors and Actuators A: Physical, Vol 73, March 9, 1999, p 138.
[71] Chris Sanders, Cronos Integrated Microsystems, Cronos is currently manufacturing a thermal actuator which is capable of delivering over 10 mN force, reference the MCNC web page: http://mems.mcnc.org/mrelact.html
[72] J. Schimkat, "Contact Materials for Microrelays", *IEEE Workshop on Micro-Electro-Mechanical Systems*, Heidelberg, Jan 25-29, 1998, p 190.
[73] Paul Slade, *Electric Contacts*, Marcel Dekker, Inc., 1999, p 449
[74] this minimum breakdown voltage is related to the work function of the contact material and the ionization energy of the atmosphere between the contacts, for gold contact surfaces and air this minimum standoff voltage is ~ 12 V.
[75] Magdaluyo, M. "Hot-swap Your Way to High Availability". *Electronic Design News*, 30 March 2000.
[76] Israelsohn, J. "Hot-swapping Power" *Electronics Design News*, 2 August 2001, pp. 54-63.
[77] Leyva, P. "Interfacing Switches and Relays to the Real World in Real Time", *Electronic Design News*, 7 June 2001.

[78] Ezekiel Kruglick and Kristofer Pister, "Micronewton Contact Characterization for MEMS Relays", *19th International Conference on Electric Contact Phenomena*, Nuremberg, September 1998, p. 53.
[79] Phipps, M. "Design and Development of Microswitches for Micro-Electromechanical Relay Matrices," Master's Thesis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, *AFIT/GE/ENG/95J-02*, 1995.
[80] Lyke, James C. and Glen A. Forman. "Space Electronics Packaging Research and Engineering", Chapter 8 of *Microengineering Aerospace Systems*, edited by Henry Helvajian, Aerospace Press, El Segundo, CA, 1999.
[81] Klein, J. "Bumped-wafer Technology Meets MOSFET Challenges", *Electronic Products*, September 2000, pp. 31-32.
[82] E.W. Dijkstra, "A Note on Two Problems in Connexion With Graphs", *Numer. Math.* **1**(1959), pp. 269-271.
[83] Alexander, M.J. and G. Robins, "New Performance-Driven FPGA Routing Algorithms", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, **15**(12):1505-1517, December 1996.
[84] Rose, J. and S. Brown, "Flexibility of Interconnection Structures for Field-Programmable Gate Arrays", *IEEE Journal of Solid-State Circuits*, **26**(3): 277-282, March 1991.
[85] R. Holm, Electrical Contacts Handbook, 2nd ed. (Spinger-Verlag, Berlin, 1958).
[86] M. Ruan, J. Shen, and C. B. Wheeler, Proc IEEE Micro Electro Mech Syst MEMS. 14, 224 (2001).
[87] X.-Q. Sun, K. R. Farmer, and W. N. Carr, in A bistable microrelay based on two-segment multimorph cantilever actuators., Heidelberg, Germany, 1998 (IEEE), p. 154.
[88] C. Angus, Brit. J. Appl. Phys. 13, 58 (1962).
[89] D. Maugis and H. M. Pollock, Acta metall. (USA). 32, 1323 (1984).
[90] A. Tonck, F. Houze, L. Boyer, J.-L. Loubet, and J.-M. Georges, J. Phys.: Condens. Matter. (UK). 3, 5195 (1991).
[91] J. Schimkat, in Contact Materials for Microrelays., Heidelberg, Germany, 1998 (IEEE), p. 190.
[92] D. Hyman and M. Mehregany, IEEE Trans. Compon. Packag. Technol. 22, 357 (1999).
[93] J. Tringe, W. Wilson, and J. Houston, Proc. SPIE. 4558, 151 (2001).
[94] M. D. Pashley, J. B. Pethica, and D. Tabor, Wear 100, 7 (1984)
[95] S. A. Joyce and J. E. Houston, Rev. Sci. Instrum. 62, 710 (1991)
[96] K. L. Johnson, K. Kendall, and A. D. Roberts, Proc. R. Soc. Lond. A, Math. Phys. Sci. (UK) 324, 301 (1971
[97] An UVOCS, Inc., Ultra-violet Ozone Cleaning System, Model T10X10/OES was used for ozonation of all substrates and tips
[98] W. E. Pendleton, A. Tackett, L. Korzeniowski, G. B. Cvijanovich, R. T. Williams, and W. C. Jones, Electr. Contacts. Proc. Annu. Holm Conf. Electr. Contacts., 109 (1998).
[99] Metals Handbook Desk Edition, 2nd ed., edited by J.R. Davis (ASM International, 1984).
[100] J. I. Pascual, J. Mendez, J. Gomez-Herrero, A. M. Baro, N. Garcia, and V. T. Binh, Phys. Rev. Lett. (USA) 71, 1852 (1993).
[101] These difficult UHV IFM measurements were performed by H. Cabibil and J. E. Houston at Sandia National Laboratories.
[102] H. B. Bakoglu. *Circuits, Interconnections, and Packaging for VLSI*. Addison-Wesley, Reading, MA 1990
[103] Drory, A. "Theory of Continuum Percolation: I. General Formalism", *Physical Review E*, **54(**6**)**:5992-6002, December 1996.
[104] Agraït, G. Rubio, and S. Vieira, Phys. Rev. Lett. (USA). 74, 3995 (1995)
[105] Davis, J.A., V. K. De, and J. D. Meindl, "A Stochastic Wire-length Distribution for Gigascale Integration (GSI): I. Derivation and Validation ", *IEEE Transactions on Electron Devices*, **45(**3**)**: 580-589, March 1996.
[106] Barabasi, A. and R. Albert, "Emergence of scaling in random networks", *Science*, 286:509-512, 1999.
[107] Cheng, Chung-Kuan; Kahng Andrew B.; and Bao Liu,"Interconnect Implications of Growth-Based Structural Models for VLSI Circuits", *Proceedings of System-Level Interconnect Prediction Workshop*, March 2001
[108] J. A. Davis, A Hierarchy of Interconnect Limits and Opportunities for Gigascale Integration, PhD Thesis, Georgia Institute of Technology, March 1999.

DISTRIBUTION LIST

DTIC/OCP
8725 John J. Kingman Rd, Suite 0944
Ft Belvoir, VA 22060-6218 1 cy

AFRL/VSIL
Kirtland AFB, NM 87117-5776 1 cy

AFRL/VSIH
Kirtland AFB, NM 87117-5776 1 cy

Official Record Copy
AFRL/VSSE, Dr. James Lyke 1 cy